A machine learning alternative to sparseness

Paloma H. Lira Fontes and Daniel O. Trad

ABSTRACT

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 traditional geophysical theory. In this work, we employ the nonlinear capabilities of ML to discern between signal and noise within the model RT space, even when conventional techniques like localization separation and smart mute fall short. This approach becomes particularly 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 U-Net's efficacy 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 dataset 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 due to shared features within the Linear RT space, compounded by spatial aliasing and the irregular geometry inherent in field data.

INTRODUCTION

The aim of processing seismic reflection data is to maintain events that are relevant for interpreters, such as the primary reflections while attenuating unwanted events referred to as noise. Multiples, a form of seismic noise, result from energy reflecting more than once and being recorded by the receivers on the surface. They are periodic in the slowness (reciprocal of velocity) domain but not in the time domain and have a larger moveout than the primary reflections, which makes it possible to separate them in certain domains, such as the Radon Transform (RT).

Ground roll is a different type of seismic noise typically found on onshore seismic data. It can usually be seen in the short offsets of the shot domain, and it has this name since it is found in the form of a cone. Different from multiple which is an event with approximately hyperbolic shape, ground roll has a linear shape which permits a different approach while separating them from the primary reflections.

The RT is a mathematical tool that maps different shaped events in the data domain into a transformed or RT domain, commonly known in geophysics as $\tau - p$, with τ representing the transformed time and p the slowness (or ray parameter). One of the advantages of applying this new domain in seismic data is that the primaries and multiples can be distinguished due to the difference in velocity and moveout. In the case of ground roll, it is due to the primaries having a different shape if compared to multiples. Consequently, the RT domain can be conveniently manipulated to mute unwanted noise while keeping the primary reflections.

Originally introduced by Johann Radon (1917), the RT has been applied in different topics such as inversion (Thorson and Claerbout, 1985), multiple attenuation (Foster and Mosher, 1992), interpolation (Trad et al., 2002), among other geophysical subjects. In seismic data processing, the RT is applied to map the events in seismic gathers by using line integrals that can map a linear, parabolic or hyperbolic-shaped event. The linear and parabolic RT are time-invariant, meaning the different basis functions are parallel with each other, so frequently they are calculated in the frequency domain. On the contrary, the hyperbolic RT is time-variant and calculated in the time domain.

Different from the inverse Fourier transform that completely restores the data, the inverse RT has some restrictions. The RT operator is not orthogonal and therefore applying the forward and inverse operators without loss of data is not trivial (Thorson and Claerbout (1985), Trad et al. (2003)). Unless the operator is orthogonal, its adjoint is not the same as its inverse. The ideal case would require unlimited data to obtain a completely invertible RT, which is not the case for seismic data. These are truncated within maximum and minimum offset and possibly missing traces thus affecting the resolution of the RT. A smooth and sparse model will both honour the data, but the gaps will be treated differently, and that is how the concept of sparse RT (Thorson and Claerbout (1985), Sacchi and Ulrych (1995), and Trad et al. (2003)) helps to address the resolution.

The RT domain has been used in the field of geophysics by some authors for noise attenuation purposes, such as by Hampson (1986) showing that the RT with parabolic basis functions is a convenient domain to filter out multiples. Foster and Mosher (1992) also discussed examples and extended multiple suppression using target-oriented parabolic RT. Trad et al. (2003) demonstrated the relevance of high-resolution sparseness in RT for higher resolution. Events shaped as a parabola or hyperbola in the data domain will be ideally represented as a point in the parabolic and hyperbolic RT domain, respectively.

A classic example of utilizing RT to mute unwanted noise (ground roll and multiples) is the case of mute using a line boundary. In this case, the fact that multiple and primary reflections are geometrically separated in the RT space since they have different velocities makes the mute as simple as a line boundary that separates the primaries on one side, where the filter value is set to one, and the noise on the other side with a filter valued as zero. In a perfect scenario signal and noise are well mapped and sampled. In reality, 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". These aliasing artifacts (Moore and Kostov, 2002) are the result of poor sampling and limited aperture on 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 Radon transform tries to address that.

To have a more accurate separation of the noise and useful events and then apply the mute the processor would need to create a boundary manually to convey the complexity of each dataset, which is difficult and time-consuming. Having this in mind, Trad (2003) introduced the idea of a mute ("smart mute") in which instead of only relying on the spatial localization of the noise (line boundary) in the RT domain, the amplitude is used as a third axis, allowing the so-called smart mute to separate overlapping events in the model space. With this approach, events that overlap in the data domain could be separated in the RT domain even if they fall into the same model parameter. This idea was used before in Harlan et al. (1984) to remove noise using focusing as the guiding decision maker. In fact, from a broad perspective, sparse transforms also use this concept while trying to enforce spatial localization by softly penalizing (muting) low-amplitude components of the inverted model. Regarding the concept of penalizing low amplitudes, it is important to notice that small amplitudes can be essential to reproduce weak events and complex waveforms. In fact, that is an issue when using a line boundary alone because visually these low amplitudes are overwhelmed by the stronger mapping from other events. Muting by amplitude alone would also suffer from the same problem.

In this work, we consider how to use the non-linearity that machine learning is able to notice to separate signal and noise from the model space when the localization separation and smart mute are not sufficient. Hence, we could consider the separation of multiples and ground roll from the RT domain even in situations where there is no full spatial separation.

As noted by Russell (2019), in the pursuit of applying physics to real-world problems, we often find ourselves simplifying complex scenarios. However, it's essential to acknowledge that the geophysical solutions generated from these simplifications often oversimplify the inherent complexities of the real world. Consequently, employing tools like machine learning can be pivotal in unveiling the unnoticed nonlinearities within these solutions, offering insights beyond the scope of traditional geophysical theory. We can relate clustering techniques to 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 data-driven 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 approach, the network tries to find patterns and predict them based on the data 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 report is to tackle challenges in traditional seismic data processing tools while applying a deep learning technique such as neural networks, more specifically U-Net. Traditional applications of RT would use line boundary mute or mute by amplitude ("smart-mute"), whereas neural networks can introduce non-linearity (inherent characteristic of real data) into this process. This enables the network to learn the crosstalk between RT spaces while predicting the mute to be done.

The core idea of this report is the application of machine learning as an alternative to traditional sparseness techniques. While sparseness works well for idealized data, it struggles with complex and overlapping events, such as the case for field data. By exploring the pixel-by-pixel (in our case window-by-window) approach the key is to make the network "intelligent" enough to handle overlapping information and other data complexities to effectively convey how this approach complements or even supersedes traditional sparseness techniques. The label choice is also really relevant for understanding the crosstalk. An important aspect of the discussion will also involve the concept of using separate channels for different types of RT spaces.

THEORY

The report Fontes et al. (2022) has a brief theoretical introduction to non-sparse and sparse Radon transforms, following the work of Thorson and Claerbout (1985), Sacchi and Ulrych (1995), Trad et al. (2003) and Trad (2001) so here we will focus on the differences between the different types of Radon transform. There is also a recollection of the U-Net architecture, a neural network.

Hyperbolic Radon transform - HRT

The hyperbolic RT, also known as velocity stack (Thorson and Claerbout, 1985), is the most suitable to map primary reflections because, in the seismic gathers, these events are hyperbola shaped. From a geometrical point of view, the HRT maps nearly hyperbolic events in the CMP gathers (data space) to points in the hyperbolic RT space by using the hyperbolic moveout equation (Yilmaz, 2001):

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

where t is the two-way travel time, τ is the zero-offset two-way travel time, x is the offset and v is the stacking velocity. Thus the HRT can be calculated by summing the amplitudes over the hyperbolas (Trad, 2001). In the discretized case Beylkin, 1987, the adjoint operator that maps from the data space to the hyperbolic RT space is given by:

$$m(v,\tau) = \sum_{x_{min}}^{x_{max}} d(x, t = \sqrt{\tau^2 + \frac{x^2}{v^2}}), \tag{2}$$

where x_{min} and x_{max} represent the offset range. Then the forward operator that maps back to the data space will be given by:

$$\tilde{d}(x,t) = \sum_{v} m(v, \tau = \sqrt{t^2 - \frac{x^2}{v^2}}). \tag{3}$$

Having the slowness (q) as the reciprocal of the stacking velocity ($v = 1/q$) equations 1, 2, and 3 can also be given in terms of p instead of v, and conveniently the hyperbolic RT plots in this report will use q in the x-axis instead of v. Because hyperbolas are time-variant curves the primaries and multiples will not be exactly parallel in the Radon domain but will reproduce a similar trend.

Linear Radon transform - LRT

The linear RT, also known as slant-stack transform (Claerbout (1985), Treitel et al. (1982)) can be applied for different goals but in the present work, we use it to separate ground roll (geometrically linear events in the shot domain) from reflections. We do that by applying the linear moveout to the data and summing over the events such as:

$$m(p, \tau) = \sum_{x_{min}}^{x_{max}} d(x, t = \tau + px), \tag{4}$$

where $p = (sin\theta)/v$ is the ray parameter or slowness, τ is the intercept time when $p = 0$ and θ is the angle between the ray being reflected and the vertical.

Parabolic Radon transform - PRT

In order to make the reflection events have a parabolic shape the idea of PRT (Hampson, 1986) was introduced taking the CMP gathers and applying NMO correction using the hyperbolic moveout (Equation 1) with the stacking velocity v of the primaries. Consequently, the primary events are flattened (straight line) and the multiples still have an approximately parabolic moveout (Yilmaz, 2001), therefore having different shapes. This will allow the summation along the parabola traveltime curve that can be represented in the discretized case by:

$$m(q, \tau) = \sum_{x_{min}}^{x_{max}} d(x, t = \tau + qx^2), \tag{5}$$

where the slowness q describes the curvature (slope) of the event (reciprocal of velocity), τ is the intersection with the zero offset and t is the time after NMO correction. While changing the value of q the basis function will match with multiples also having a strong signature in the parabolic RT domain. Since those events have different curvatures (velocities) it is possible to map them separately. Low values of q allow the mapping of flattened events (reflections in the CMP domain after NMO correction) whereas higher values of q would map multiples.

There is also the pseudo-hyperbolic RT (Foster and Mosher, 1992), which is a different way of decomposing hyperbolic events in the time-invariant basis functions (Trad, 2001). The summation will be along the hyperbola but it will depend on the offset and a specific depth, such as:

$$m(q, t) = \sum_{x_{min}}^{x_{max}} d(x, t = \tau + q[\sqrt{z_{zef}^2 + x^2} - z_{ref}]), \tag{6}$$

where z_{zef} is the specific depth.

Hybrid Radon transform

In the case of land seismic data, ground roll is a geometrically linear shaped event and reflections can be approximated to parabolic shaped events. Then Trad (2001) and Trad

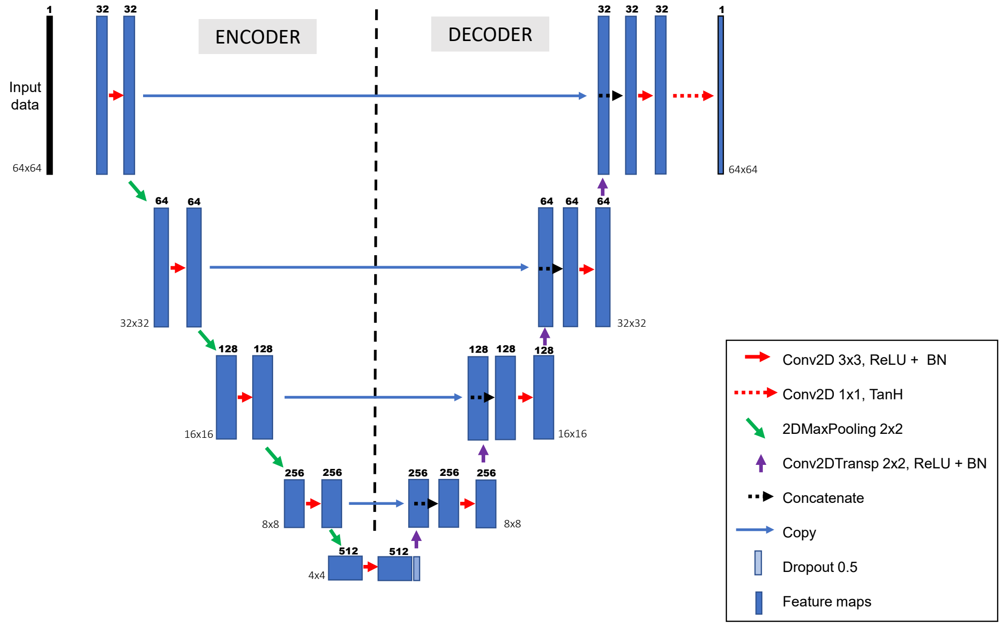

FIG. 1. Schematic representation of the modified U-Net architecture (Fontes et al., 2022).

et al. (2001) introduced the concept of the hybrid RT. It is based on the fact that linear (equation 4) and parabolic (equation 5) RT are time-invariant and therefore have similar axes in the RT domain as well as similar parameters, such as τ, q and x. Because of the time invariance, the transform can be calculated independently for every frequency, which is also computationally attractive since it does lots of small operations instead of a big one in the whole time domain Because it imposes sparseness in the q coordinate and time the RT has high quality with very few artifacts.

U-net: convolutional neural network

One of the most applied CNNs architectures is the U-Net (Ronneberger et al., 2015), mostly used for image segmentation problems. It can be applied in different fields, such as cell detection in microscope image (Falk et al., 2019), salt interpretation in seismic data (Zeng et al., 2019) among others.

The U-Net architecture is characterized by its assembly of convolutional and pooling layers within an encoder-decoder framework. Figure 1 illustrates the U-net structure employed in this report, consistent with previous reports Fontes et al. (2022) and Trad (2022). 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. It also updates the weights by concatenating them with its corresponding encoder outputs, forming an interconnected U-shaped structure. The choice of label holds significant importance within this method as they contain critical information used by the network to learn how to identify specific features within an image.

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. Another way to update the weights is by using inference. By starting with weights from a previous train the network will not start addressing the weights randomly, instead, it will take advantage of previous knowledge and try to deepen the understanding of the problem towards improving the predictions. To evaluate the model performance is usually used a cost function or loss, such as the mean square error – MSE. The loss will show the amount of error the network is having while predicting by measuring the distance between the prediction and the target vector. It is formulated as (Géron, 2019):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}^{(i)}) - \mathbf{y}^{(i)})^2, \tag{7}$$

where, $\mathbf{x}$ is the vector of predictions and $\mathbf{y}$ is vector of the true model (label). And so, if the MSE decreases during the training then the predictions will be as close to the label (model of the true multiples) as possible.

EXAMPLES AND DISCUSSION

As previously mentioned the U-Net is usually applied in classification problems, but in the present work, we modified that and used this network to perform regression to predict the desired events (multiples and attenuation of ground roll) in the RT domain. All of the U-Net architecture and hyperparameters such as the number of filters, their weights and widths, strides and padding type explained in the previous section were kept the same throughout all the examples. The same architecture was used in the report Fontes et al. (2022) since the tests presented here are a continuation of the work shown there.

To better understand how this U-Net works 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 real seismic data are handled while predicting the desired events (multiples for future adaptive subtraction, or predicting the ground roll attenuated version of the dataset). This report also aims to understand how different RT spaces talk with each other, a relationship named crosstalk.

Bridge with Parabolic and Hyperbolic RT - prediction of multiples:

To train the U-Net (Figure 1), we used Madagascar to generate (employing convolutional model) 120 synthetic shot gathers (380 receivers, shot interval 10 meters, total record time 4.5 seconds with 0.004 seconds of temporal sampling and 1125 samples per trace) from the same 8 geological layers velocity model used on our previous report (Fontes et al., 2022). Shot 98 can be seen in Figure 2 input data (multiples and primaries) and input label (multiples only) for training. Then the Hyperbolic RT (Figure 2 (a)), sparse Hyperbolic RT (Figure 2 (b)) and Parabolic (Figure 2 (c)) inverse operator (data to model space) is applied, generating RT panels. In the RT panels we can see primaries and multiples are spatially separated, with primaries aligned to the left and multiples to the right since they have higher travel times. Then these RT images (1125 (τ) x 200 (p^2)) passed through some data preparation steps (normalization and scaling) and by a windowing process (compartmentalized into smaller patches of 64x64) these are then the data inputted into the U-Net, which will be assigned the task of training and predicting multiples.

It is important to note that some primaries and multiples are almost or totally overlapping in the shot domain and therefore will be really close in the HR domain, so that is something to keep close attention to if the network will understand and map this difference.

The output of the network will be the 578 RT panels (1125x200) with, hopefully, the predicted multiples. The last step of the workflow is to apply the RT (forward operator) using Least Squares to reconstitute the data from the RT domain back to the CMP domain (and back to the shot domain).

Some observations can be made about Figure 2:

- The choice of label is really important for this application of U-Net. Since the prediction is based on the training the choice of the label will set the tone for how the prediction will look like.
- The sparse version does not necessarily provide a better prediction. 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. A consequence of this is the so-called butterfly effect (Marfurt et al., 1996). 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.
- The HRT (Figure 2 (a)), compared with the sparse HRT (Figure 2 (b)) does a better job on the prediction on the long offsets portion of the predicted shot. However, the sparse HRT does a better control of the background noise. Yet, the PRT (Figure 2 (c)) seems to have overall a better prediction of multiple, since it contains a good balance between maintaining the long offsets (even though stretch mute is being applied) and having less background noise.

Different from last year's report (Fontes et al., 2022), where we discussed the possibilities of using various RT panels as different channels, now the idea is to try to see if Hyperbolic and Parabolic RT (time variant and invariant, respectively) can complement each other. The trouble is 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, we employed some further tests.

Figure 3 illustrates the workflow for the Bridge 1 test. By having the input data as multiples and primaries in the PRT and labels as multiples and primaries HRT, we can predict an "intermediary1" output, which does not have physical meaning since the input and the label have different x-axis. This intermediary output bridges into being the input RT panel, and by using the HRT as a label we are able to predict the RT with multiples

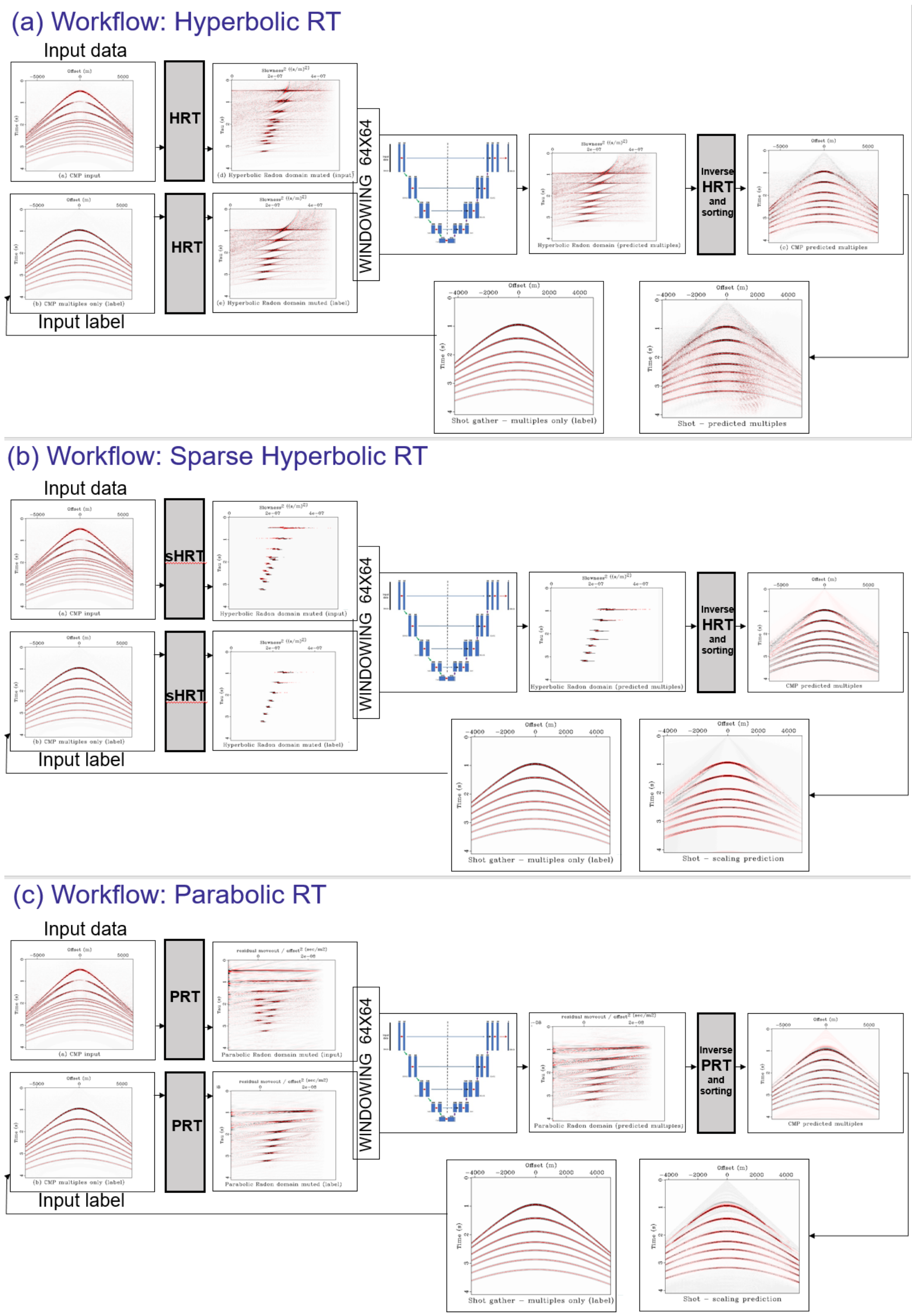

FIG. 2. Workflow for (a) Hyperbolic RT, (b) sparse Hyperbolic RT and, (c) Parabolic RT. All of them have an input shot with primaries and multiples and, the input label is the shot with multiples only.

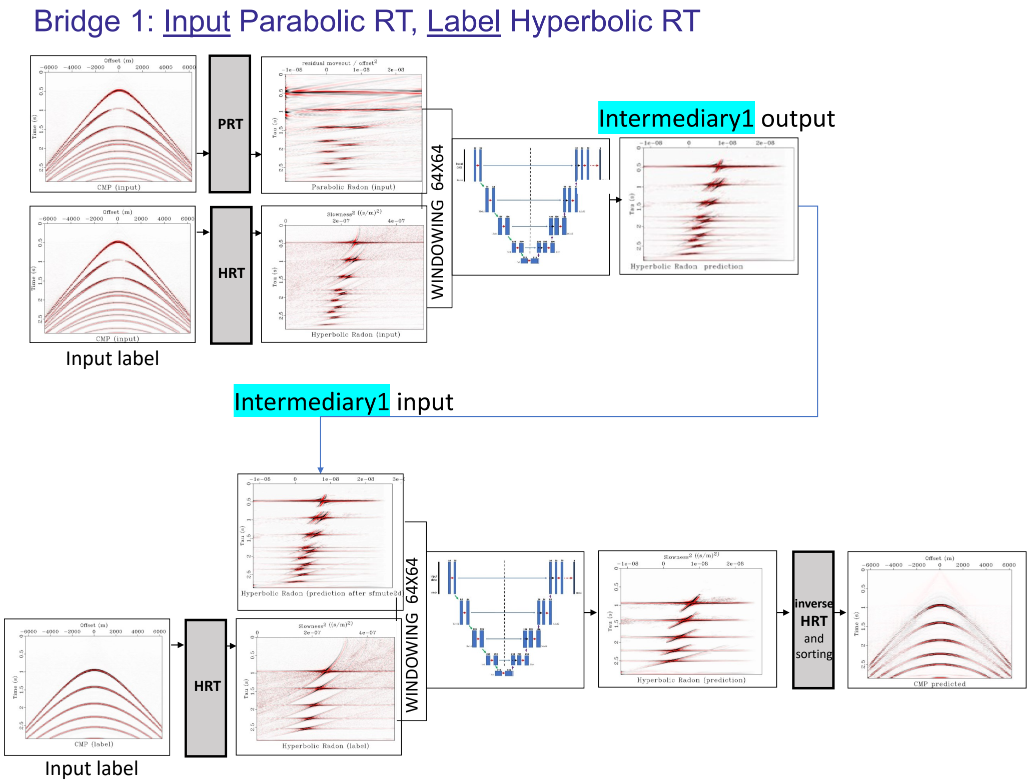

FIG. 3. Bridge 1: a workflow that uses Hyperbolic RT as label

only and converted back to the data domain. Figure 4 shows the workflow for Bridge 2, which has a similar idea to Bridge 1 but now we are trying to understand if it is better to use the PRT as the label. By comparing the results of the prediction of both of the bridges the predicted CMP that looks the most like its label is Bridge 2.

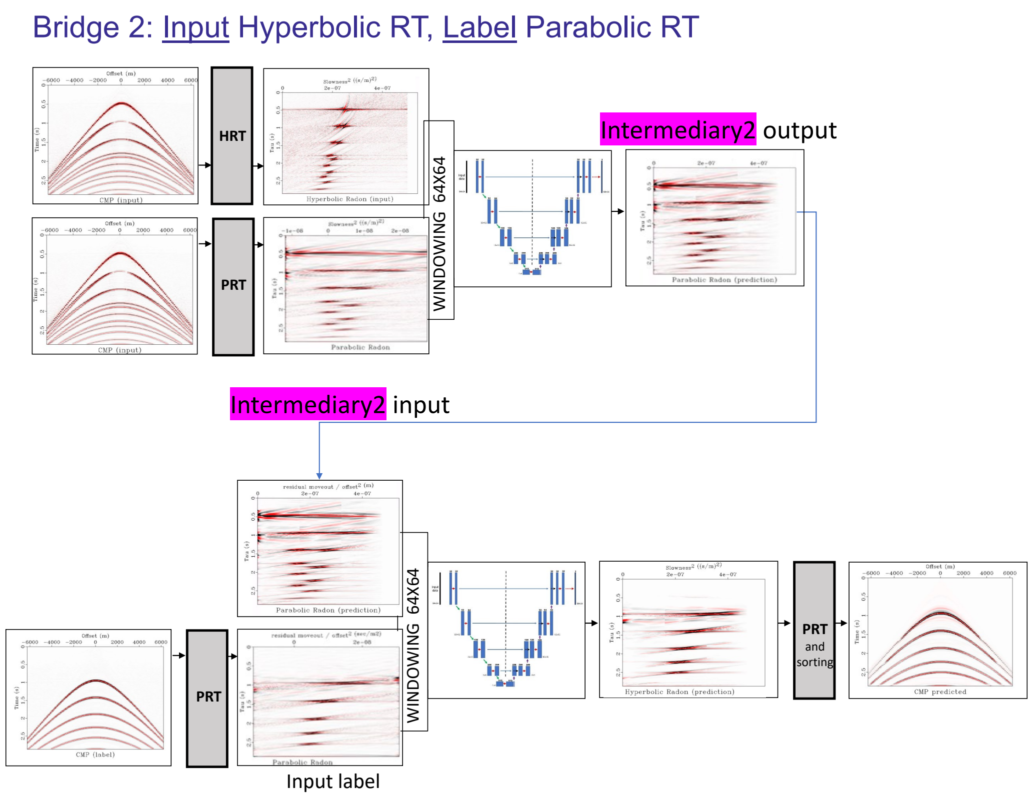

FIG. 4. Bridge 2: a workflow that uses Parabolic RT as label

Hybrid RT – predicting ground roll attenuation:

Usually, seismic data processors do not visualize the RT space (Figure 5 (b, b') and, Figure 6 (b, b')) while applying RT. What is usually done is a check in the reconstruction of the data (Figure 5 (c, c') and, Figure 6 (c, c')), but it can be helpful to have a visual of seismic events mapping into different regions of the RT domain. With the aim of understanding whether it is possible to map different events into different RT spaces (Trad et al., 2001) used 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, which is the case for ground roll and reflections since they are mapped in the data space into lines and approximately parabolic shapes, respectively, in the RT domain.

The linear RT (left side of Figure 5 (b) and (b')) maps reflections and ground roll from the data space into ellipses and linear events in the model space, respectively. In the case of the parabolic RT 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 we can see in Figure 5 (b') if compared with when the ground roll is not present (5 (b)). As well as the band-limited nature of seismic data and irregularity in geometry which is usually the case for field data. One of the advantages of working with synthetic data is we can organize it in a regular grid whereas field data (Figure 5 (a, a')) have irregularities, such as variable offsets, among others.

The Spring Coulee 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 (field data usually has irregular parameters, for example: 623, 629, 652), 0.002 s sample rate and 2001 number of samples per trace. Because in this report we are trying to tackle real data issues we generated the synthetic data (Figure 5 a and a') using the same geometry and approximate velocity model of the mentioned field data.

Through understanding that different seismic events can be mapped into different basis functions one possibility is to reconstruct the data while only selecting a region of the RT that maps the basis function of your choice (Figure 7). For instance, one can restrict the Linear RT x-axis (Figure 7 (b')) and only select an area that has ellipses (reflections) and reconstruct the data using only the desired operator by taking a smaller portion of the dq (slowness) axis (Trad et al., 2001). This would technically map back mainly the reflections (Figure 7 (c')). However in field data things are not that linear, and this kind of uncontrolled filtering can also lead to the loss of useful signal as we can see in Figure 7 (c).

There are several methods for ground roll removal in literature and this report does not aim to suggest a new method. Rather, we would like to discuss how we could apply the Hybrid RT in 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 can work as 3 (2 linear and 1 parabolic) and 2 (1 linear and 1 parabolic) channels

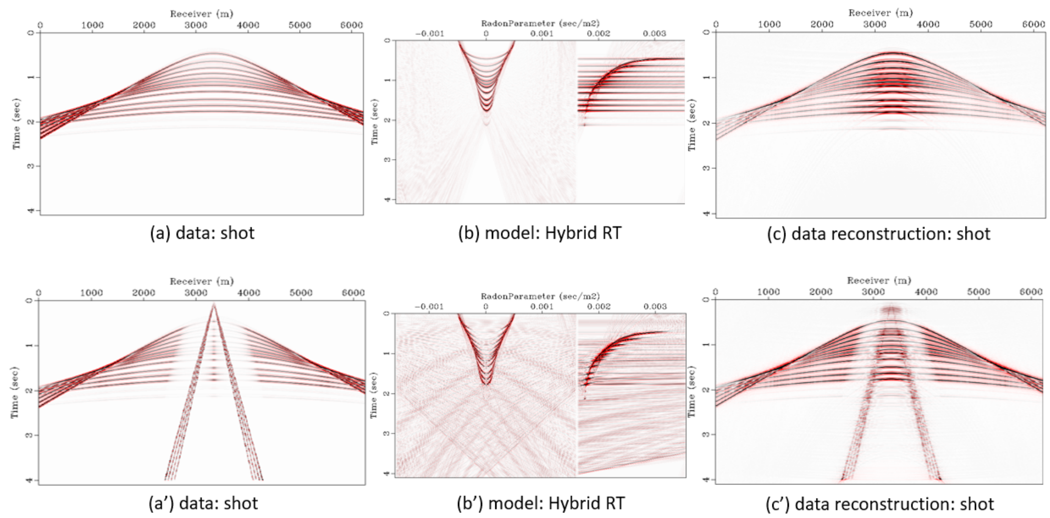

FIG. 5. Synthetic shot (using the convolutional model) with reflections only (a) and with ground roll (a') using velocity model and geometry from Spring Coulee field data; Hybrid (Linear and Parabolic) RT of reflections only (b) and its reconstruction (c). Hybrid RT of reflections and ground roll (b') and its reconstruction (c').

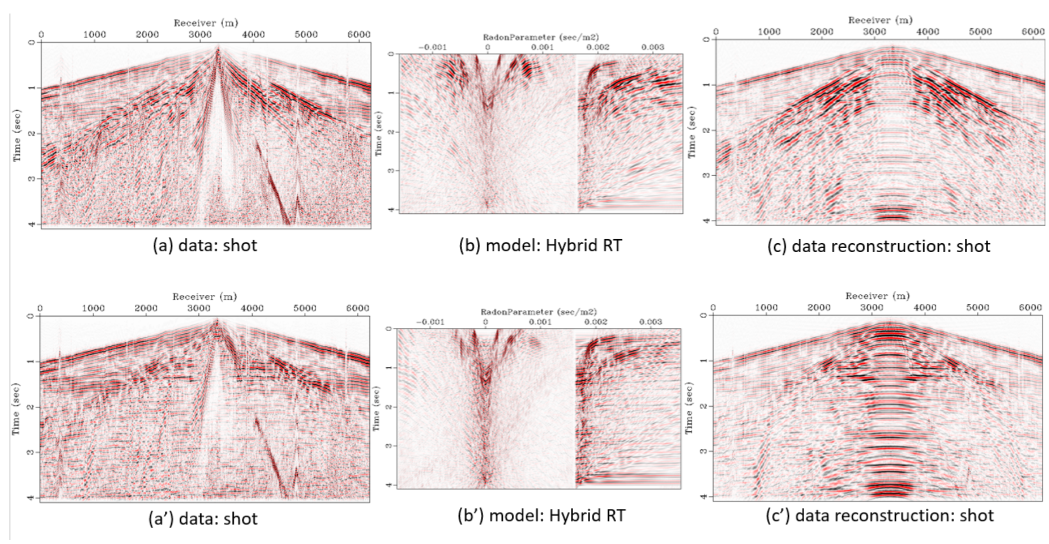

FIG. 6. (a) original Spring Coulee shot with visible ground roll and some reflections, its (b) Hybrid (Linear and Parabolic) RT and its (c) shot reconstruction. (a') Spring Coulee original shot after being FK filtered in Vista with more visible reflections, its (b') Hybrid (Linear and Parabolic) RT and its (c') shot reconstruction.

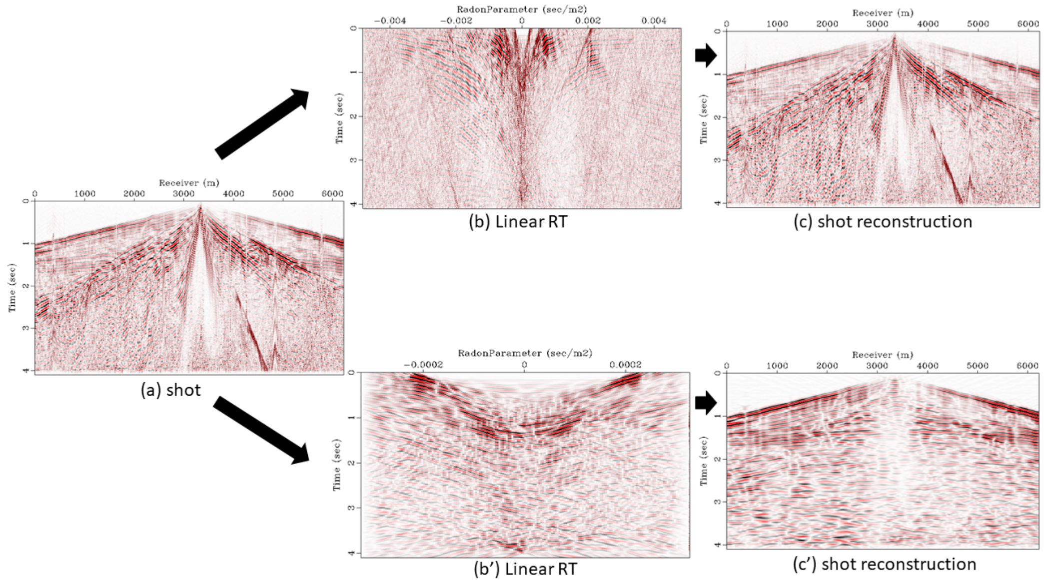

FIG. 7. (a) original Spring Coulee shot, (b) Linear RT with a wide x-axis, and its (c) shot reconstruction. (b') Linear RT with a restricted x-axis that focuses on the ellipses, and its (c') shot reconstruction.

in the U-net training. We would also like to 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) we would like to keep.

Test I: one channel

The Hybrid RT - 1 channel workflow is shown in Figure 8. This test was done using original data, meaning that there are all events in it, as the input and the FK filtered of the input as the label, mimicking as there were only reflections in there. This last observation introduces uncertainty for the prediction since creating a label purely with reflections in real data is not possible. Another limitation is the reconstruction of the Hybrid RT (Figure 8 (g, *, **)) is not the best since we are using only the adjoint operator and not Least Squares (with external iterations).

The prediction of the U-Net (Figure 8 (g)) has some similarities with the prediction of the case with $\sim$ reflections only (Figure 8 (**)), which it shows that the ML is doing a fair prediction work. Also comparing the Radon panels from the label (Figure 8 (d)) and the prediction (Figure 8 (f)) are really similar, with little connection with the ground roll aliasing nature.

Test II: three channels

After evaluating that the U-Net could do a fair prediction with the simple case of 1 channel we then tried to further investigate which of the 3 channels would give the better prediction: Linear, negative and positive since the data has split-spread geometry, and Parabolic. In order to do that we split the Hybrid RT (Figure 8 (c)) panel into three: $-$ Linear (Figure 9 (a)), $+$ Linear (Figure 9 (b)) and Parabolic (Figure 9 (c)) RT. Each one of

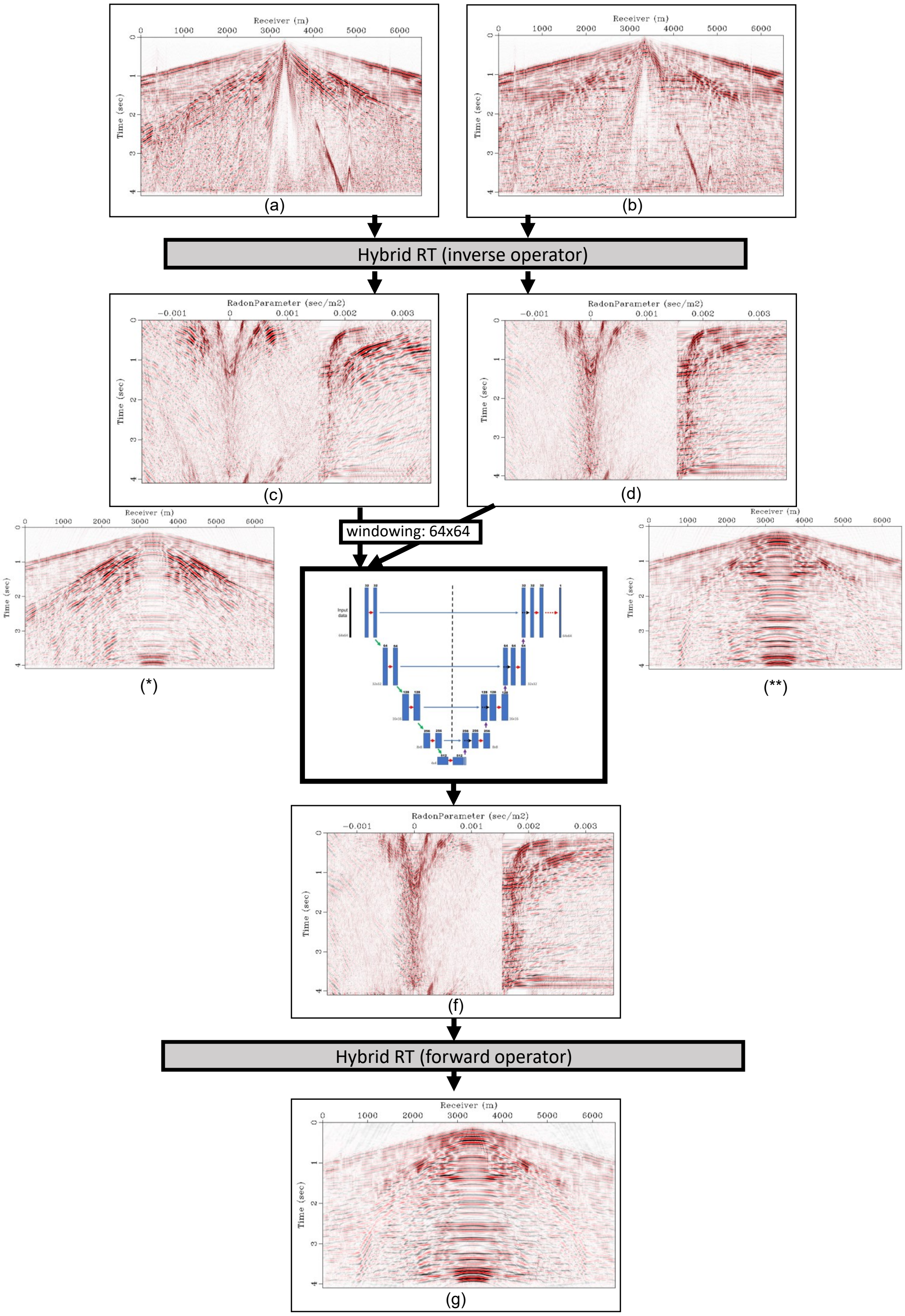

FIG. 8. Hybrid RT one channel workflow: Spring Coulee shot gathers are the input data (a), with the ground roll and primaries and, the input label (b), with the original being FK filtered in Vista to approximate a label $\sim$ reflections only. Then the Hybrid RT (inverse operator) is applied to generate the hybrid panels of the input (c) and label (d) to feed the U-Net (e). The network then predicts, after training, the hybrid panels with $\sim$ reflections only (f), attenuating the ground roll. The inverse Hybrid RT (forward operator) using the adjoint operator is then applied to return the data to the shot domain (g).

the three portions of the Hybrid RT will be the input for the U-Net but in order to do that some adjustments needed to be taken in the x-axis (slowness) to make sure that all 3 panels had the same size since the ML process is done pixel by pixel.

The prediction of the U-Net using either $-$ Linear, $+$ Linear or Parabolic as a label is shown in Figure 11 (c, d, e). We can see that the one that looks the closest to the reconstruction of the Linear (positive and negative have similar results, Figure 11 (c) and (d), respectively). The Parabolic as label seems to be the one with the most leakage of ground roll and another kind of noise in the prediction (Figure 11 (e)).

Test III: two channels

While using 3 channels we separate the Linear RT into positive and negative. Since they both had the best prediction reconstruction we decided to then test the prediction using only 2 channels: Linear and Parabolic. Because of the pixel-by-pixel ML approach in order to make this test possible, some adjustments were needed to make sure that all 2 panels, Linear and Parabolic, had the same size, as shown in Figure 10.

The prediction of the U-Net using either Linear or Parabolic as a label is shown in Figure 11 (f, g), respectively. Figure 11(f) has a better prediction than (c) and (d) since now the whole Linear RT space is being used to reconstruct the data, which also seems to have the one better. Comparing the reconstruction of the Parabolic as the label (Figure 11(e) and (g)) seems to show that the 2 channels prediction has a better reconstruction. This could be justified by the increase in the size of the slowness axis from the 3 channels to the 2 channels case.

The last comparison that should be done is all the predictions with the label of the real data (Figure 6 (a')). The shot reconstruction (Figure 11 (c, d, e, f, g)) that has the most reflections in fact the 3 channels Parabolic as a label. This can be due to the fact that most of the ground roll lives in the Linear RT space, and therefore when using the Parabolic as the label the reconstruction targets the concentration of energy in the reflections more than in the ground roll. Although it is important to mention that there is information that is not reflections that lives in the Parabolic space, and because we are dealing with field data it will never be possible to truly and fully separate data and ground roll without losing some of the reflection information since this information also lives in the Linear RT space.

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 Parabolic as label. The choice of label is really important for this application of U-Net. However, it is not clear 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. Train with Hybrid RT split into three channels, -Linear, +Linear and Parabolic, and using Parabolic RT labels resulted in better 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.

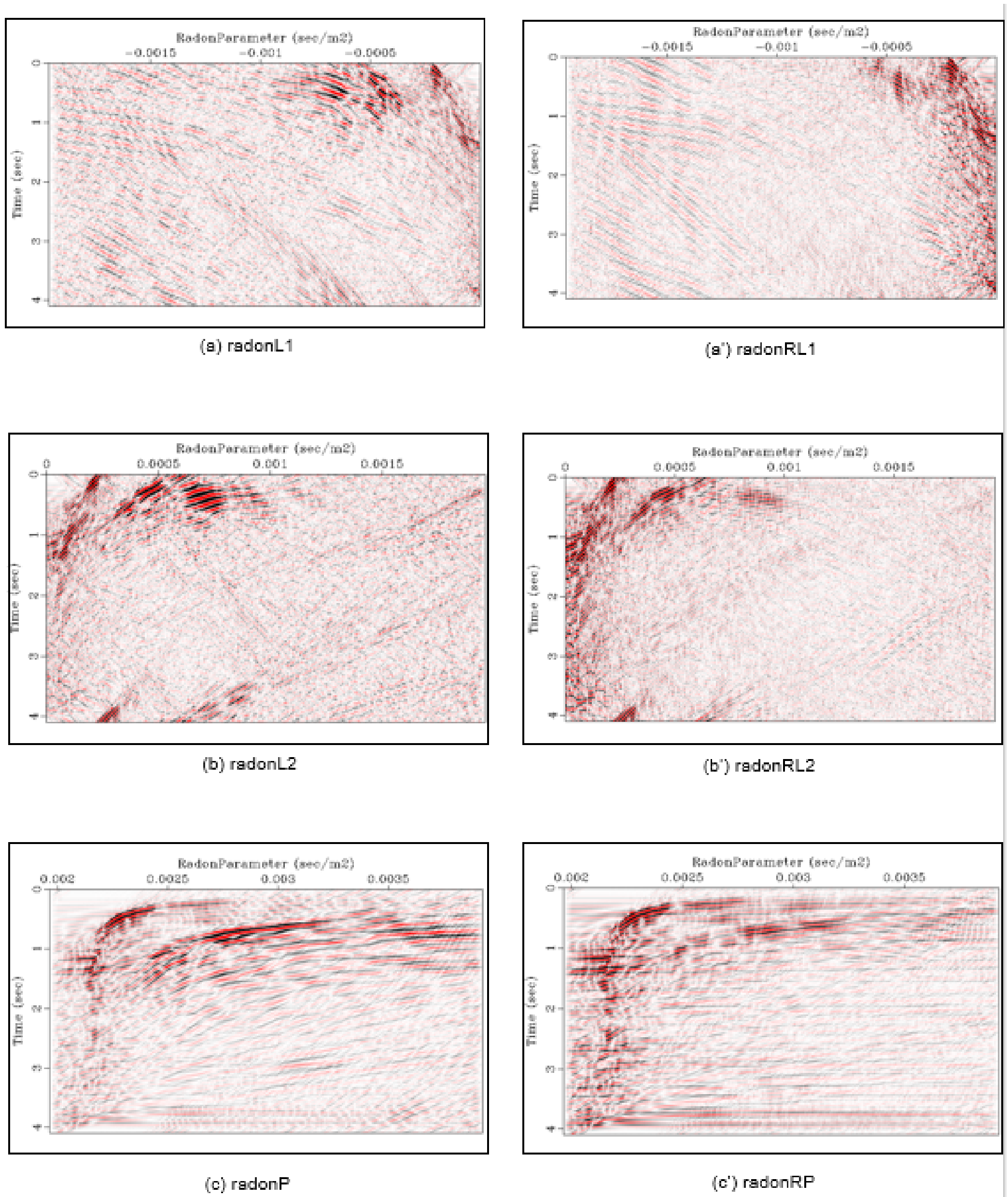

FIG. 9. The input Spring Coulee Hybrid RT was split into 3 channels: (a) $-$ Linear, (b) $+$ Linear, and (c) Parabolic RT panels of ground roll, reflections and every other event from a field data. The same split was done with its label:(a') $-$ Linear, (b') $+$ Linear, and (c') Parabolic RT panels of $\sim$ reflections only (FK filtered version of the input).

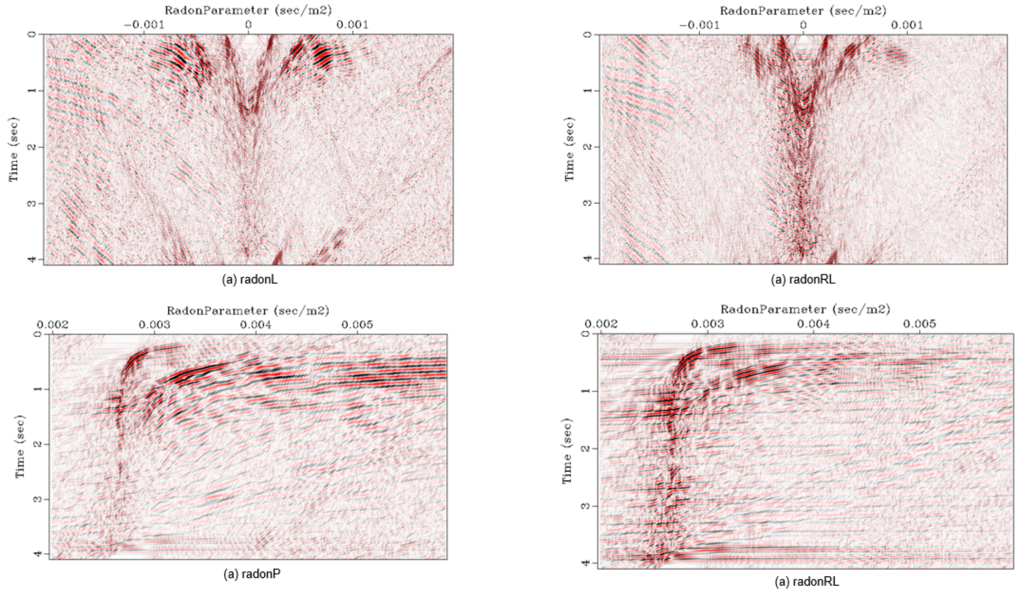

FIG. 10. The input Spring Coulee Hybrid RT was split into 2 channels: (a) Linear, and (b) Parabolic RT panels of ground roll, reflections and every other event from field data. The same split was done with its label: (a') Linear, and (b') Parabolic RT panels of $\sim$ reflections only (FK filtered version of the input).

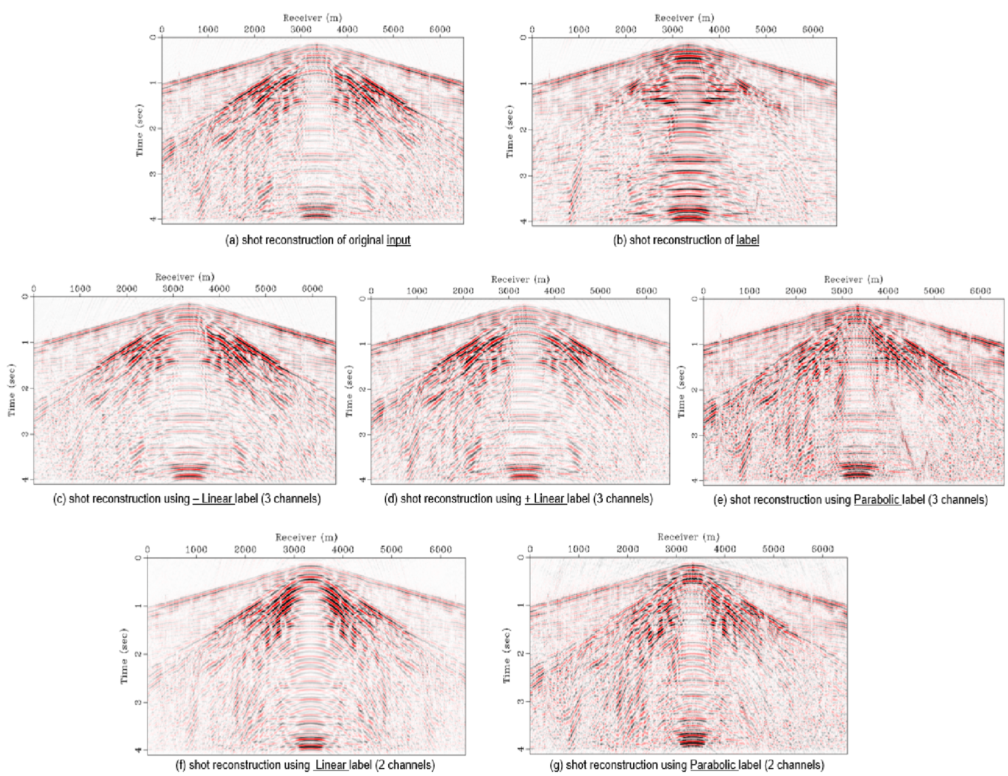

FIG. 11. (a) Spring Coulee dataset: the adjoint operator of the Hybrid RT input (a), and from its $\sim$reflections only (FK filtered from the original) (b). 3 channels: (c) reconstruction of the prediction using $-$Linear RT as the label, (d) reconstruction of the prediction using $+$Linear RT as label and, (e) reconstruction of the prediction using Parabolic RT as the label. 2 channels: (f) reconstruction of the prediction using Linear RT as label and, (g) using Parabolic as the label.

ACKNOWLEDGMENTS

The sponsors of CREWES are gratefully thanked for continued support. This work was funded by CREWES industrial sponsors, NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 543578-19. One of the authors of this report was supported by the CSEG Foundation. We also thank Kai Zhuang and Ivan Sanchez for their discussions, brainstorming about the topics covered in this report and help with Madagascar.

REFERENCES

Beylkin, G., 1987, Discrete radon transform: IEEE transactions on acoustics, speech, and signal processing, **35**, No. 2, 162–172.

Claerbout, J. F., 1985, Imaging the earth's interior, vol. 1: Blackwell scientific publications Oxford.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K. et al., 2019, U-net: deep learning for cell counting, detection, and morphometry: Nature methods, **16**, No. 1, 67–70.

Fontes, P. H. L., Trad, D. O., and Sánchez-Galvis, I. J., 2022, The use of u-net and radon transforms for multiple attenuation: CREWES Research Report, **34**.

Foster, D. J., and Mosher, C. C., 1992, Suppression of multiple reflections using the radon transform: Geophysics, **57**, No. 3, 386–395.

Géron, A., 2019, Hands-on machine learning with Scikit-Learn, Keras, and Tensor-Flow: O'Reilly Media, Inc.

Hampson, D., 1986, Inverse velocity stacking for multiple elimination: SEG Technical Program Expanded Abstracts, 422–424.

Harlan, W. S., Claerbout, J. F., and Rocca, F., 1984, Signal/noise separation and velocity estimation: Geophysics, **49**, No. 11, 1869–1880.

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, cp–5.

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.

Russell, B., 2019, Machine learning and geophysical inversion—a numerical study: The Leading Edge, **38**, No. 7, 512–519.

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, Society of Exploration Geophysicists, 5686–5690.

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

Trad, D. O., 2001, Implementations and applications of the sparse radon transform: Ph.D. thesis, The University of British Columbia.

Trad, D. O., 2003, Interpolation and multiple attenuation with migration operators: Geophysics, **68**, No. 6, 2043–2054.

Trad, D. O., 2022, Combining classical processing with deep learning: CREWES Research Report, **34**.

Trad, D. O., Sacchi, M. D., and Ulrych, T. J., 2001, A hybrid linear-hyperbolic radon transform: Journal of Seismic Exploration, **9**, No. 4, 303–318.

Trad, D. O., Ulrych, T., and Sacchi, M., 2003, Latest views of the sparse radon transform: Geophysics, **68**, No. 1, 386–399.

Trad, D. O., Ulrych, T. J., and Sacchi, M. D., 2002, Accurate interpolation with high-resolution time-variant radon transforms: Geophysics, **67**, No. 2, 644–656.

Treitel, S., Gutowski, P., and Wagner, D., 1982, Plane-wave decomposition of seismograms: Geophysics, **47**, No. 10, 1375–1401.

Yilmaz, Ö., 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: Society of exploration geophysicists.

Zeng, Y., Jiang, K., and Chen, J., 2019, Automatic seismic salt interpretation with deep convolutional neural networks, *in* Proceedings of the 2019 3rd international conference on information system and data mining, 16–20.