

# Deep Unsupervised 4D Seismic 3D Time-Shift Estimation with Convolutional Neural Networks

Jesper Søren Dramsch<sup>✉</sup>, Anders Nymark Christensen<sup>✉</sup>, Mikael Luthje<sup>✉</sup>

**Abstract**—We present a novel 3D warping technique for the estimation of 4D seismic time-shift. This unsupervised method provides a diffeomorphic 3D time shift field that includes uncertainties, therefore it does not need prior time-shift data to be trained. This results in a widely applicable method in time-lapse seismic data analysis. We explore the generalization of the method to unseen data both in the same geological setting and in a different field, where the generalization error stays constant and within an acceptable range across test cases. We further explore upsampling of the warp field from a smaller network to decrease computational cost and see some deterioration of the warp field quality as a result.

**Index Terms**—4D seismic, time-lapse, deep learning, unsupervised learning, 3D time-shift, neural network

## I. INTRODUCTION

SEISMIC time-lapse data consists of two 3D reflection amplitude cubes that represent the subsurface they were collected from. These cubes are acquired years apart with expected changes in the subsurface due to e.g. hydrocarbon production. The differences in the subsurface cause changes in both amplitudes and velocities, which introduces misalignment of seismic reflectors. Measuring the misalignment and aligning these surfaces to obtain a reliable difference cube is one of the main disciplines in 4D seismic processing.

These time shifts are most commonly obtained by windowed cross-correlation and other statistic or signal processing approaches [1]. Considering the recent advances of machine learning in imaging and domain transfer, we explore possibilities of alignment with convolutional neural networks. Machine learning approaches, however, most commonly require labeled data to find a mapping  $f(x) = y$ , with  $x$  being the input data,  $f$  being the blackbox algorithm like a neural network, and  $y$  being the labels or target.

A common problem in machine learning for subsurface science is determining the ground truth. Obtaining information from the subsurface is often prohibited by cost, and e.g. core samples are highly localised data and is often altered by the extraction method as well as the sheer act of unearthing the sample. Additionally, synthetic data may introduce the inverse crime [2] of using the same theory to generate and invert data. Luckily, the physics of medical imaging and inversion is very similar to geophysics, where methods can be validated and fine-tuned. The main method discussed in this paper is adapted from the medical imaging literature.

The lack of ground truths leads to another problem that deep learning address but do not solve. For classic neural networks, we need to know a target label dataset. In the case of this problem formulation this means knowing a prior warp velocity.

In 4D seismic this would mean employing an established method to obtain time shifts. This would effectively result in abstracting that method in a neural network, or modelling the warp, which would lead to committing the inverse crime. Logically, this lead us to explore unsupervised methods.

We discuss several options for architectures for mapping the monitor seismic cube to the base seismic cube directly within the network. This is possible in unsupervised configurations but depending on the architecture of the network this problem can be ill-constrained and generate non-physical mappings. One warranted criticism of deep learning and neural networks is the lack of explainability and limited interpretability. However, we employ a deep neural network to obtain warp velocity vectors, a 3D equivalent of time shifts, for dense deterministic warping instead of directly obtaining the warped result from a neural network. This enables us to interpret the warping vectors and constrain the warp path in addition to the warp result.

Moreover, we present the first 4D seismic 3D Timeshift estimator with uncertainty measures. We achieve this by implementing a variational layer that samples from a Gaussian with the reparametrization trick (cite). Therefore, we can counteract some of the influence of noise on the performance of the network.

## II. THEORY

Extracting time shifts from 4D seismic data is often done trace-wise, which limits the problem to depth. This provides sufficient results for simple problems. However, geologically complex systems and pre-stack time shifts benefit from obtaining 3D time-shifts. We discuss classical 3D time-shift extraction methods, we then go on to discuss relevant deep learning methods. These methods extract time-shifts with different constraints we explore. For brevity we present the results of the best method to date, developed for the medical domain: VoxelMorph.

The goal of both conventional and machine learning methods is to obtain a warp velocity field  $\mathbf{u}(x, y, z)$  that ideally aligns two 3D cubes  $B$  and  $M$  within given constraints. That means a sample  $m[x, y, z]$  will be aligned by adjusting  $m[x+u_x, y+u_y, z+u_z]$ . In image processing this is considered "dense alignment" or "dense warping", hence we need a dense vector field to align each sample in the base and the monitor cube. Generally,  $\mathbf{u}(x, y, z) \in \mathbb{R}^3$ , which implies interpolation to obtain the warped result.

### A. Conventional Methods

Most conventional methods in 4D seismic warping focus on 1D methods [3], which include local 1D cross-correlation, dynamic time warping [4], optical flow methods and methods based on Taylor expansion [5]. We do not cover these methods in detail, but focus on the limited applications of 3D methods in 4D seismic warping.

1) *Local 3D Cross Correlation*: [6] introduced local 3D cross-correlation as a method for image alignment. This method relies on calculating the multi-dimensional cross-correlation windowed by a Gaussian with a specified radius. The correlation results are normalized to avoid spurious correlations by amplitude fluctuations and high-amplitude events. Subsequently the cross-correlation result is searched for peaks using the following triple sum:

$$c[u_x, u_y, u_z] = \sum_{x, y, z = -\infty}^{\infty} b[x, y, z] \cdot m[x + u_x, y + u_y, z + u_z],$$

with  $c$  being the cross-correlation lag. The computational complexity of this method is  $\mathcal{O}(N_s \times N_l)$  with  $N_s$  being the total number of samples and  $N_l$  being the total number of lags.

Stabilization of the results of 3D cross-correlation is obtained by applying spectral whitening of the signals and smoothing the images with a Gaussian filter without increasing the computational complexity despite the windowing function [6].

2) *Inversion-based methods*: [7] describe a non-linear inversion approach, with the objective function being

$$\mathbb{E} = |\mathbf{d} - f(\mathbf{m})|^2 + |\nabla_x(\mathbf{m})|^2 + |\nabla_y(\mathbf{m})|^2 + |\nabla_z^2(\mathbf{m})|^2 \quad (1)$$

with  $\mathbf{m}$  being the model vector,  $\mathbf{d}$  being the data vector. The non-linear inversion is constrained by applying the first-derivative to the spatial dimensions  $z$ ,  $y$  and Laplacian in  $z$  to obtain a smooth solution. [8] implement a geostatistical joint inversion that uses the geostatistical information combined with data constraints as a prior in a Bayesian inversion scheme.

$$P(x|geostats, data) \propto \exp\left(-\left(x - \mu\right) \cdot C^{-1} \cdot \frac{x - \mu}{2}\right) \quad (2)$$

with  $C$  being the posterior covariance matrix and  $\mu$  being the posterior mean.

### B. Machine Learning Methods

The machine learning methods discussed in this section are imaging based, and therefore rely on recent advances of convolutional neural networks (CNN) in deep learning. We discuss different approaches that include supervised and unsupervised / self-supervised methods. These methods are all based on convolutional neural networks (CNNs).

CNNs are a type of neural network that is particularly suited to imaging approaches. They learn arbitrary data-dependent filters that are optimized based on the chosen objective via gradient descent. These filters can operate on real images, medical images, or seismic data alike. The convolutional filter

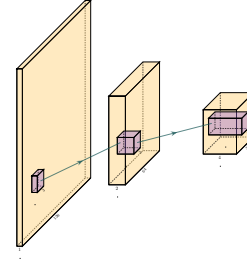


Fig. 1. Schematic convolutional neural network

benefits from weight sharing, making the operation efficient and particularly suited to GPUs or specialized hardware. In Figure 1 we show a schematic image, that is convolved with moving 3x3 filters repeatedly to obtain a spatially down-sampled representation. These convolutional layers in neural networks can be arranged in different architectures that we explore in the following analysis of prior methods in image alignment.

1) *Supervised CNNs*: Supervised end-to-end CNNs rely on reliable ground truth, including the time shifts being available. Training a supervised machine learning system requires both a data vector  $x$  and a target vector  $y$  to train the blackbox system  $f(x) \Rightarrow y$ . This means that we have to provide extracted time-shifts from other methods, which implicitly introduce assumptions from that method into the supervised model. Alternatively, expensive synthetic models would be required.

The supervised methods are largely based on Optical Flow methods [9], [10]. The FlowNet [9] architecture is based on an Encoder-Decoder CNN architecture. Particularly, FlowNet has reached wide reception and several modifications were implemented, namely FlowNet 2.0 [11] improving accuracy, and LiteFlowNet [12] reducing computational cost. SpyNet [10] and PWC-Net [13] implement stacked coarse-to-fine networks for residual flow correction. PatchBatch [14] and deep discrete flow [15] implement Siamese Networks [16] to estimate optical flow. Alternatively, DeepFlow [17] attempts to extract large displacements optical flow using pyramids of SIFT features. These methods introduce varying types of network architectures, optimizations, and losses that attempt to solve the optical flow problem in computer vision.

2) *Unsupervised CNNs*: Unsupervised or self-supervised CNNs only rely on the data, relaxing the necessity for ground truth time shifts. [18] reformulates the FlowNet architecture into an unsupervised optical flow estimator with bidirectional census loss. The UnFlow network relies on the smooth estimation of the forward and backward loss, then adds a consistency loss between the forward and backward loss and finally warps the monitor to the base image to obtain the final data loss. Considering that optical flow has historically underperformed on seismic data, due to both smoothness and illumination constraints. However, UnFlow replaces the commonly used illumination loss by a ternary census loss [19] with the  $\epsilon$ -modification by [20]. While this bears possible promise for seismic data, UnFlow implements 2D losses as opposed to a 3D implementation we focus on.

### 3) Cycle-consistent Generative Adversarial Networks:

Cycle-GANs are an unsupervised implementation of Generative Adversarial Networks that are known for domain adaptation [21]. These implement two GAN networks that perform a forward and backward operation that implements a cycle-consistent loss in addition to the GAN loss. The warping problem can be reformulated as a domain adaptation problem. This implements two Generator networks  $F$  and  $G$  and the according discriminators  $D_X$  and  $D_Y$ . These perform a mapping  $G : X \rightarrow Y$  and  $F : Y \rightarrow X$ , trained via the GAN discrimination. The cycle-consistency implements  $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$  with the backwards cycle-consistency being  $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ .

Cycle-GANs such as pix2pix [22] separate image data into a content vector and a texture vector, which could bear promise in the seismic domain, adapting a wavelet vector and an interval vector [23]. However, the confounding of imaging effects, changing underlying geology, changing acquisition, etc makes the separation non-unique. Moreover, extracting the time shift information and conditioning the GAN is a very complex problem. The Recycle-GAN [24] addresses temporal continuity in videos, this is however hard to transfer to seismic data, considering the low number of time-steps in a 4D seismic survey as opposed to videos. Furthermore, the lack of interpretability of GANs at the point of writing, prohibits GANs from replacing many physics-based approaches, like the extraction of time-shifts.

## III. METHOD

The Voxelmorph [25] implements a U-net [26] architecture to obtain a dense warp velocity field and subsequently warps the monitor cube to match the base cube. A U-net architecture is particularly suited for segmentation tasks and transformations with smaller than usual amounts of data. The short-cut concatenation between the input and output layers stabilizes training and avoid the vanishing gradient problem. It is particularly suited to stable training in this image matching architecture. In Figure 2 the U-Net is the left-most stack of layers, arranged in an hourglass architecture with shortcuts. These feed into a variational layer  $\mathcal{N}(\mu, \sigma)$ , the variational layer is sampled with the reparameterization trick, due to the sampler not being differentiable TKCITE. The resulting differential flow is integrated using the VecInt layer from Neuron TKCite and passed into a spatial transformation layer. This layer transforms the monitor cube according to the warp velocity field obtained from the integrated sampler. The result is used to calculate the data loss between the warped image and the base cube.

More formally, we define two 3D images  $\mathbf{b}, \mathbf{m}$  being the base and monitor seismic respectively. We try to find a deformation field  $\phi$  parameterized by the latent variable  $z$  such that  $\phi_z : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ . The deformation field itself is defined by this ordinary differential equation (ODE):

$$\frac{\partial \phi^{(t)}}{\partial t} = v(\phi^{(t)}), \quad (3)$$

where  $t$  is time,  $v$  is the stationary velocity and the following holds true  $\phi^{(0)} = \mathbf{I}$ . The integration of  $v$  over  $t = [0, 1]$

provides  $\phi^{(1)}$ . This integration represents and implements the one-parameter diffeomorphism in this network architecture. The variational Voxelmorph formulation assumes an approximate posterior probability  $q_\psi(z|\mathbf{b}; \mathbf{m})$ , with  $\psi$  representing the parameterization. This posterior is modeled as a multivariate normal distribution with the covariance  $\Sigma_{z|m,b}$  being diagonal:

$$q_\psi(z|\mathbf{b}; \mathbf{m}) = \mathcal{N}(z, \mu_{z|m,b}, \Sigma_{z|m,b}), \quad (4)$$

the effects of this assumption are explored in [27].

The approximate posterior probability  $q_\psi$  is used to obtain the variational lower bound of the model evidence by minimizing the KL divergence with  $p(z|\mathbf{b}; \mathbf{m})$  being the intractable posterior probability. Following the full derivation in [27], considering the sampling of  $z_k \sim q_\psi(z|\mathbf{b}; \mathbf{m})$  for each image pair  $(\mathbf{b}, \mathbf{m})$ , we compute  $\mathbf{m} \circ \phi_{z_k}$  the warped image we obtain the loss:

$$\begin{aligned} \mathcal{L}(\psi; \mathbf{b}, \mathbf{m}) &= -\mathbb{E}_q[\log p(\mathbf{b}|z; \mathbf{m})] \\ &\quad + \mathbf{KL}[q_\psi(z|\mathbf{b}; \mathbf{m}) || p_\psi(z|\mathbf{b}; \mathbf{m})] \\ &\quad + \text{const} \\ &= \frac{1}{2\sigma^2 K} \sum_k \|\mathbf{b} - \mathbf{m} \circ \phi_{z_k}\|^2 \\ &\quad + \frac{1}{2} [\text{tr}(\lambda D \Sigma_{z|x;y}) - \log \Sigma_{z|x;y}] \\ &\quad + \mu_{z|m,b}^T \Lambda_z \mu_{z|m,b} + \text{const}, \end{aligned} \quad (5)$$

we can sample  $\mu_{z|m,b}$  and  $\Sigma_{z|m,b}$  as variational layers in a neural network. Given the diagonal constraint on  $\Sigma$ , we define the variational layer as the according standard deviation  $\sigma$  of the corresponding dimension. Therefore we sample  $\mathcal{X} \sim \mathcal{N}(\mu, \sigma^2)$  using the reparameterization trick first implemented in variational auto-encoders [28]

Defining the architecture and losses as presented in [27], ensures several benefits. The registration of two images is domain-agnostic, which enables us to apply the medical algorithm to seismic data. The warp field is diffeomorphic, which ensures physically viable, topology-preserving warp velocity fields. Diffeomorphisms have recently gained great attention in the medical field, particularly with large deformation diffeomorphic metric mapping (LDDMM) TKCITE, which is computationally expensive and has therefore not found great use in the wider field of geophysics, due to larger amounts of data. Moreover, this method implements a variational formulation based on the covariance of the flow field. 3D warping with uncertainty measure has not been used in seismic data before.

The network is implemented using Tensorflow [29] and Keras [30] our implementation is based on the original code in the Voxelmorph package [31].

## IV. EXPERIMENTS & RESULTS

### A. Experimental Setup

The experimental setup for this paper is based on a variation of the modified Voxelmorph [25] formulation. We extended the network to accept patches of data, because our seismic cubes are generally larger than the medical brain scans and therefore

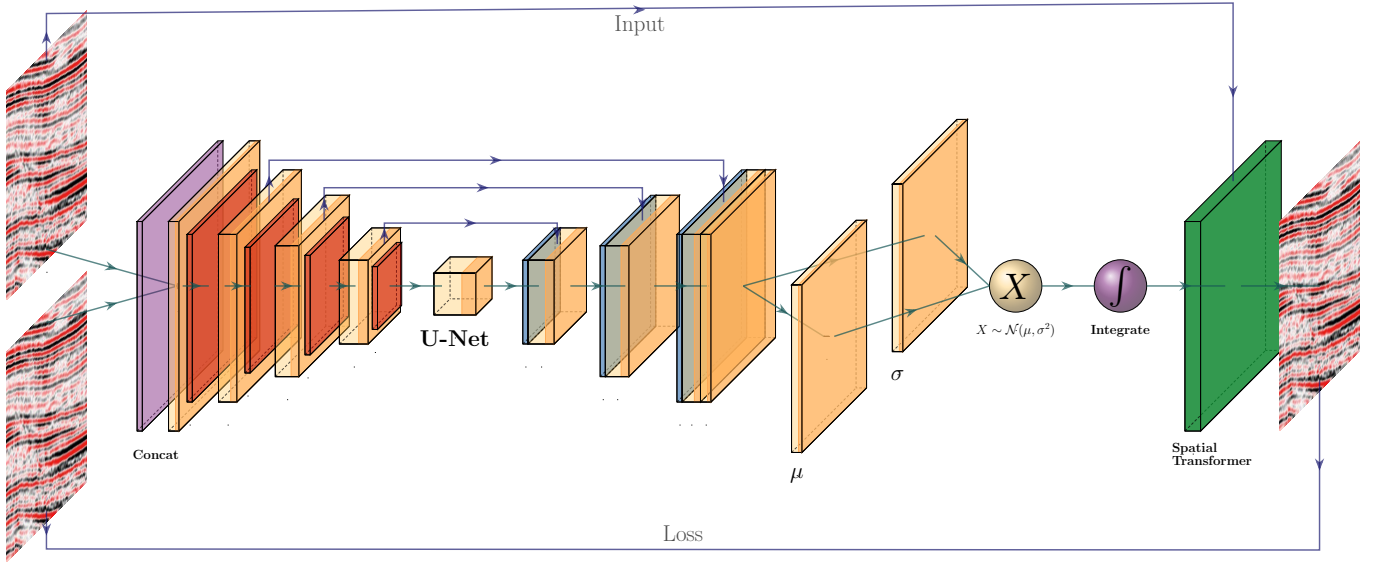


Fig. 2. Voxelmorph architecture

exceed the memory limits of our GPUs. Moreover, Voxelmorph in its original formulation provides sub-sampled flow fields, this is due to computational constraints. We decided to modify the network to provide full-scale flow fields, despite the computational cost. This enables direct interpretation of the warp field, which is common in 4D seismic analysis. However, we do provide an analysis in part IV-B3 of the sub-sampled flow-field interpolated to full scale, in the way it would be passed to the Spatial Transformer layer.

The network definition for the subsampled flow field differs from the definition in Figure 2 that the last upsampling and convolution layer in the Unet, including the skip connection, right before the variational layers  $(\mu, \sigma)$  is omitted. That leaves the flow field at a subsampled map by a factor of two. Computationally, this lowers the cost on the Integration operation before resampling for the Spatial Transformer.

The data situation for this experiment is special in the sense that the method is self-supervised. We therefore do not provide a validation dataset during training. While we would be content with the method working on the field data by itself, we do validate the results on separate data from the same field which was acquired with different acquisition parameters and at different times. Moreover, we test the data on seismic data from an adjacent field that was acquired independently. The field geology and therefore seismic responses are very different. Due to lack of availability we do not test the trained network on land data or data from different parts of the world. Considering, that the training set is one 4D seismic monitor-base pair, a more robust network would emerge from training on a variety of different cubes.

### B. Results

The network presented generates warp fields in three dimensions as well as uncertainty measures. In Figure 4 we show the results on the data the unsupervised method was trained

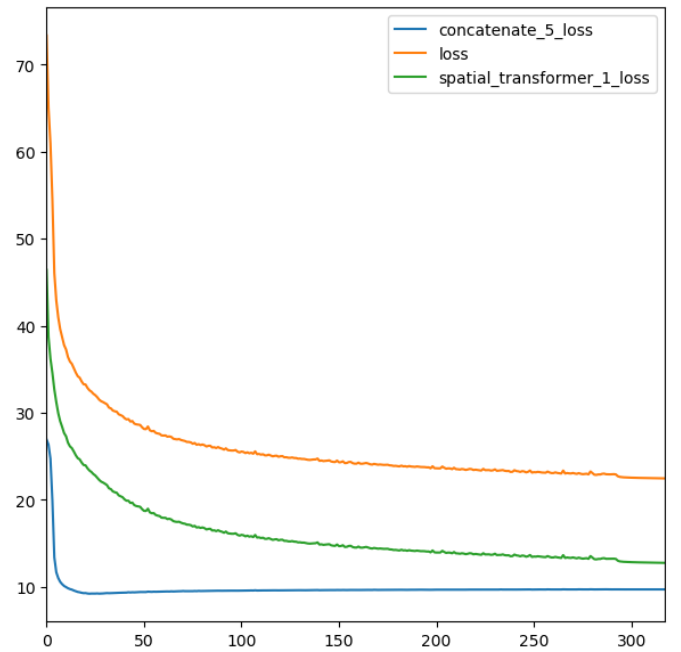


Fig. 3. Training Losses over time.

on. While in supervised methods, this is not a viable results, we pose that in addition to generalizability of the method, performance on the data itself is a good result. Considering the network is trained to find an optimum warp field for the data it was originally trained on, we go on to test the network on data from the same field, that was recorded with significantly different acquisition parameters in Figure 7. These results test the networks generalizability on co-located data, therefore not expecting vastly differing seismic responses from the subsurface itself. The are imaging differences and differences in equipment in addition to the 4D difference however. In



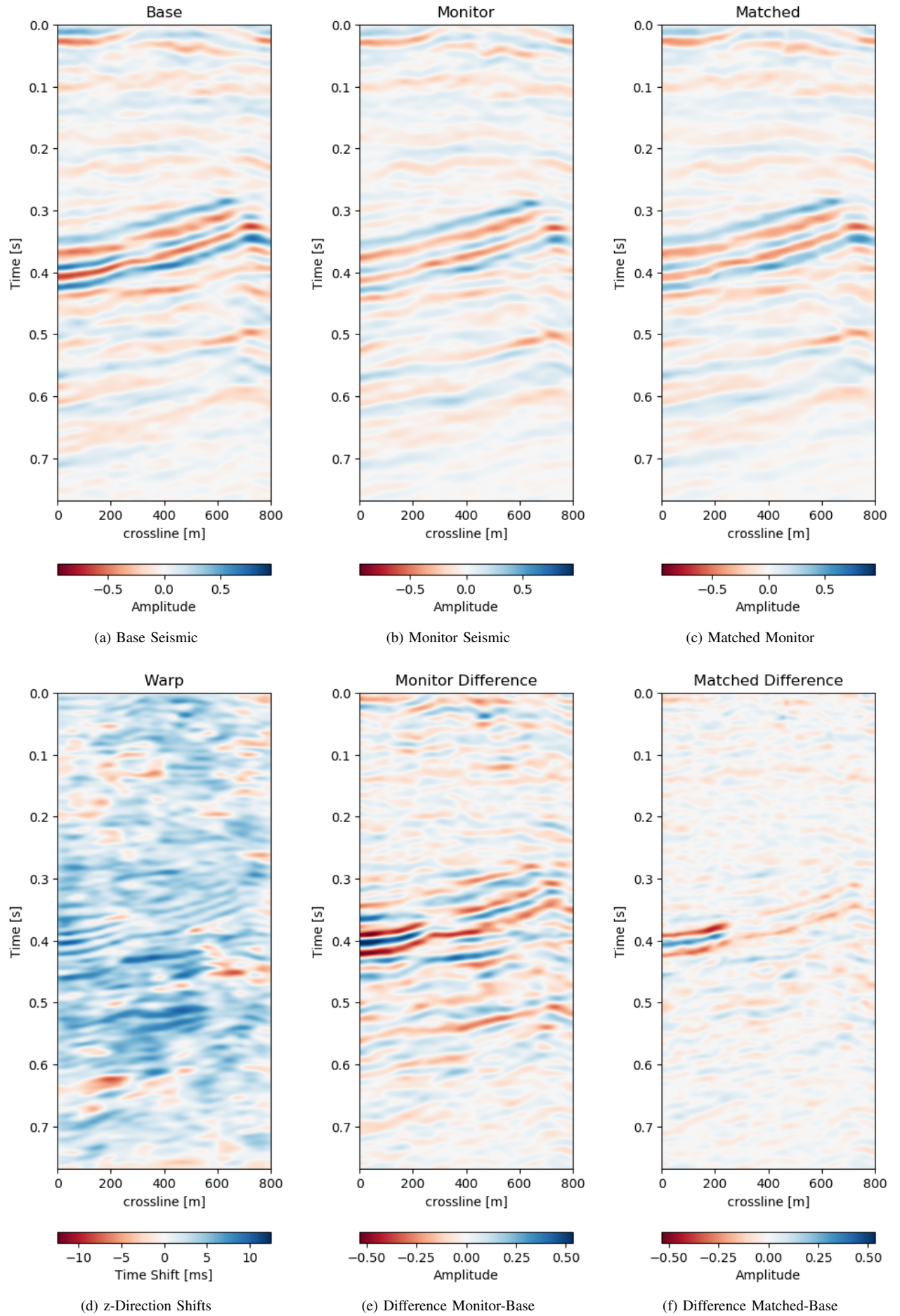


Fig. 4. Comparison of seismic data before and after matching. And the result for the amplitude difference in these data.

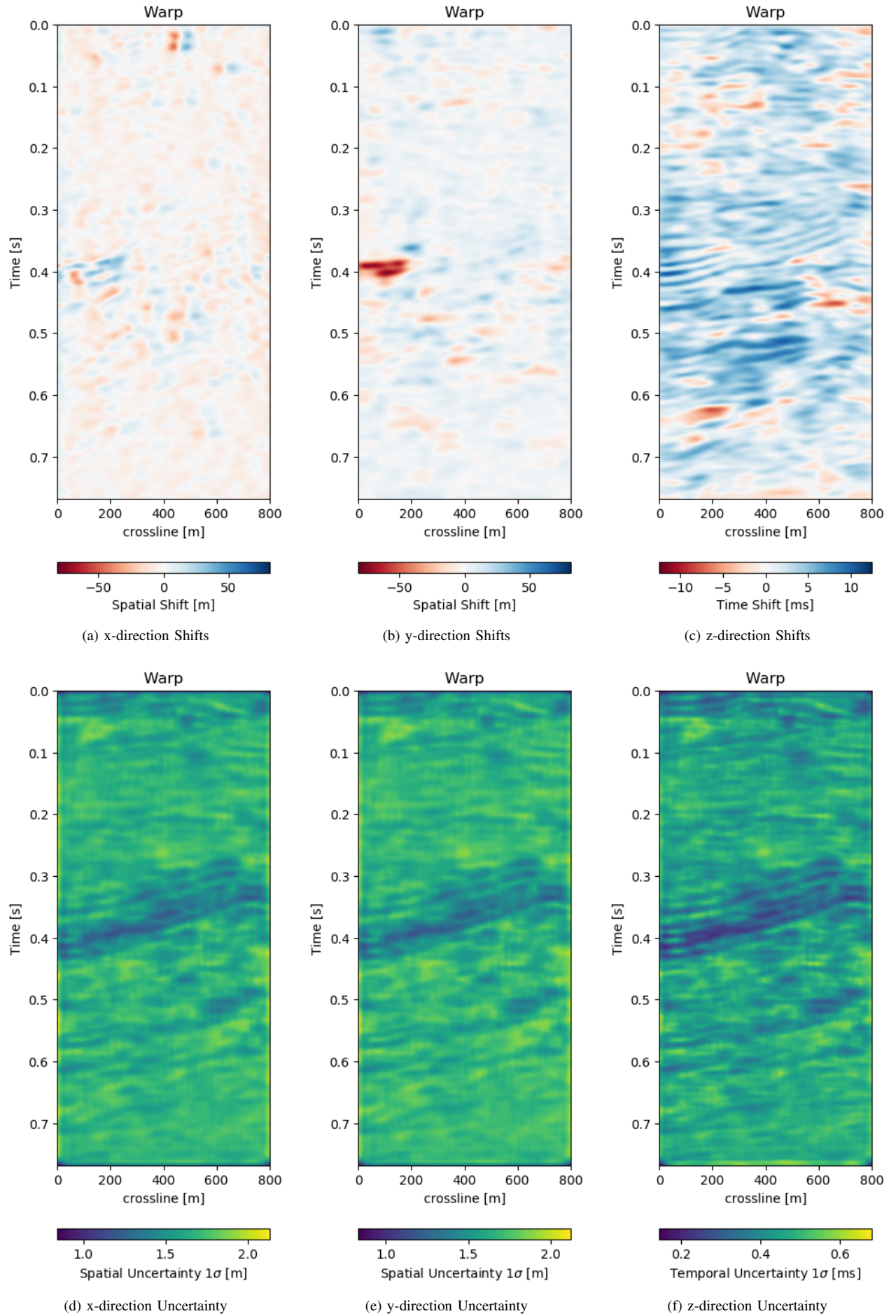


Fig. 5. Warp field with uncertainty that accompanies Figure 4

Run	Monitor RMS	Matched RMS	Ratio %	Monitor MAE	Matched MAE	Ratio %
Train	0.1047	0.0525	50.1	0.0744	0.0348	46.7
Test A	0.0381	0.0237	62.2	0.0291	0.0172	59.1
Test B	0.0583	0.0361	62.0	0.0451	0.0254	56.4

TABLE I

QUANTITATIVE EVALUATION OF RESULTS. TRAINING RECALL, TEST A - SAME FIELD, DIFFERENT ACQUISITION, TEST B - DIFFERENT FIELD, DIFFERENT ACQUISITION

Figure 9 we use the network on unseen data from a different field. The geometry of the field, as well as the acquisition parameters are different, making generalization a challenge.

In Figure 4 we collect six 2D panels from the 3D warping operation. In Figure 4a and Figure 4b we show the unaltered base and monitor respectively. The difference between the unaltered cubes is shown in Figure ???. In Figure 4c we show the warped result by applying the z-warp field in Figure 4d, as well as the warp fields in (x,y) direction fully displayed in Figure 5 including their respective uncertainties. The difference of the warped result in Figure ??? is calculated from the matched monitor in Figure 4c and the base in Figure 4a.

It is apparent that the matched monitor significantly reduced noise by mis-aligned reflections. In Table I we present the numeric results. These were computed on the 3D cube for an accurate representation. We present the root mean square (RMS) and mean absolute error (MAE) and the according difference between Monitor and Matched Difference results. We present RMS and MAE to make the values comparable in magnitude as opposed the mean squared error (MSE). We present both values, because the RMS value is more sensitive to large values, while MAE scales the error linearly therefore not masking low amplitude mis-alignments. Both measurements show a reduction on the train data to 50% or below. The test on both the validation data on the same field and the test data on another field show a similar reduction, while the absolute error differs in a stable manner.

In Figure 5 we present the three dimensional warp field to accompany the results in Figure 4. Figure 5a, 5b, and 5c show the warp field in x, y, and z-direction. The z-direction is generally referred to as time shifts in 4D seismic. Figure 5d, 5e, and 5f contain the corresponding uncertainties in x, y, and z-direction obtained from the network.

1) *Recall to Training Data:* In Figure 4 we evaluate the results of the self-supervised method on the training data itself. The main focus is on the main reflector in the center of the panels. The difference in Figure ??? shows that the packet of reflectors is out of alignment, causing a large difference, which is corrected for in Figure ???. One point of concern when implementing the diffeomorphic constraint are discontinuities like faults. The topmost section in the panel of Figure 4c shows the alignment of a faulted segment in the monitor to an unfaulted segment in the base. The fault appearing is most likely due to vastly improved acquisition technology for the monitor.

The warp fields in Figure 5 are an integral part in QC-ing the validity of the results. Physically, we expect the strongest changes in the z-direction in Figure 5c. The changes in Figure 5a and Figure 5b show mostly sub-sampling magnitude

shifts, except for the x-direction shifts around the fault in the top-most panel present in the monitor in Figure 4b. Figure 5a and Figure 5b show strong shifts at 0.4s on the left of the panel which corresponds to the strong amplitude changes in the base and monitor. On the one side these correspond to the strongest difference section, additionally these are geological hinges, which are under large geomechanical strain. However, these are very close to the sides of the warp, which may cause artifacts. Figure 5d, Figure 5e, and Figure 5f show the uncertainty of the network. These uncertainties are across the bank within the 10% range of the sampling rate. The certainty within the bulk package in the center of the panels is the lowest in x-, y-, and z-direction. While being relatively lower in the problematic regions discussed before.

The warp field in Figure 4d contains some reflector shaped warp vectors around 0.4 s, which is due to the wavelet mismatch of the 1988 base to the 2005 monitor. The diffeomorphic nature of the network aligns the reflectors in the image, which causes some reflector artifacts in the z-direction maps.

2) *Generalization of the Network:* While the performance of the method on a data set by itself is good, obtaining a trained model that can be applied on other similar data sets is pejorative even to self-supervised methods. We test the network on two test sets, Test A is conducted on the same geology with unseen data from a different acquisition, while Test B is on a different field and a different acquisition. The network was trained on a single acquisition relation (2005a - 1988). In Figure 7 we present the crossline data from the same field the network was trained on. The data sets was however acquired at a different calendar times (2005b - 2012), with different acquisition parameters. It follows that although the geology and therefore the reflection geometry is similar, the wavelet and hence the seismic response are vastly different. This becomes apparent when comparing the base Figure 7a to Figure 4b, which were acquired in the same year.

The quantitative results in Table I for Test A generally show lower absolute errors compared to the training results in Part IV-B1. The reduction of the overall amplitudes in the difference maps is reduce by 40%. The unaligned monitor difference in Figure 7e shows a strong coherent difference around below the main packet of reflectors around 0.3 s to 0.4 s. This would suggest a velocity draw-down in this packet. While the top half of the unaligned difference contains some misalignment, we would expect the warp field to display a shift around 0.35 s, which can be observed in Figure 7d. The aligned difference in Figure 7f contains less coherent differences. The difference does still show some overall noise in the maps. This could be improved upon by a more diverse training set. The higher resolution data from 2005 and 2012 possibly has an influence on the result too. Regardless, we can see some persisting amplitude difference around 0.4 s which appears to be signal as opposed to some misalignment noise above. The warp fields in Figure 8 show relatively smooth warp fields in x- and y-direction. The warp field in Figure 8f shows overall good coherence, including the change around 0.4 s we would expect. The uncertainty values are in sub-sampling range, with the strongest certainty within the strong reflector packet at 0.35 s.

Test B shows a very similar reduction of overall errors in Table I. The RMS is reduced by 38% and the MAE is reduced more slightly more in comparison to Test A. In Figure 9 we present the seismic panels to accompany Test B. The data in Figure 9a and Figure 9b is well resolved and shows good coherence. However, the unaligned difference in Figure 9e shows very strong variations in the difference maps. Figure 9f reduces these errors significantly, bringing out coherent differences in the main reflector at 0.27 s and around 0.5 s. Above the reflector packet at 0.2 s we can see strong chaotic differences in Figure 9e, due to the faulted nature of the geology. The network aligns these faulted blocks relatively well, however, some possible artifacts persist. This is consistent with the warp fields in Figure 10. The x- and y-direction in Figure 10d and Figure 10e respectively show overall smooth changes, around faults, these changes are stronger. The z-direction changes are consistent with the Training validation and Test A, where the changes are overall stronger. This is also consistent with our geological intuition.

3) *Subsampled Flow*: The original Voxelmorph implementation uses a subsampled warp field. This has two benefits, namely a smoother warp velocity field and reduced computational cost. The aforementioned results were obtained using a full-scale network. In Figure 6 we present the full scale and upsampled results on the training set. The matched difference in Figure 6b contains more overall noise compared to Figure 6a. This is congruent with the warp fields in the figure. The upsampled z-direction warp field in Figure 6d seems to have some aliasing on the diagonal reflector around 0.4 s. This explains some of the artifacts in the difference in Figure 6b. The overall warp velocity in Figure 6d is smoother compared to the full-scale field. However, the general structure of coherent negative and positive areas matches in both warp fields, while the details differ. The main persistent difference of the reflector packet at 0.4 s seems similar, nevertheless, the differences further up slope to the right are smoother in the full scale network result and have stronger residual amplitudes in the upsampled network.

## V. CONCLUSION

We introduce a deep learning based self-supervised 4D seismic warping method. We evaluate our network on the training data and two different independent test sets. We do not expect the aligned difference to be exactly zero, due to actual physical changes in the imaged subsurface. Although the network is unsupervised, a transfer to unseen data is desirable and despite some increase in the overall error possible. The warping on the training data is very good and the warp fields are coherent and reflect the physical reality one would expect. The transfer to unseen data works well, although the misalignment error increases. The decrease in both RMS and MAE is consistent across test sets.

Furthermore, we implement a variational scheme which provides uncertainty measures for the time shifts. On the data presented, we obtain subsample scale uncertainties across all directions. The main assumption of the network is a diffeomorphic deformation, which is topology preserving. We show

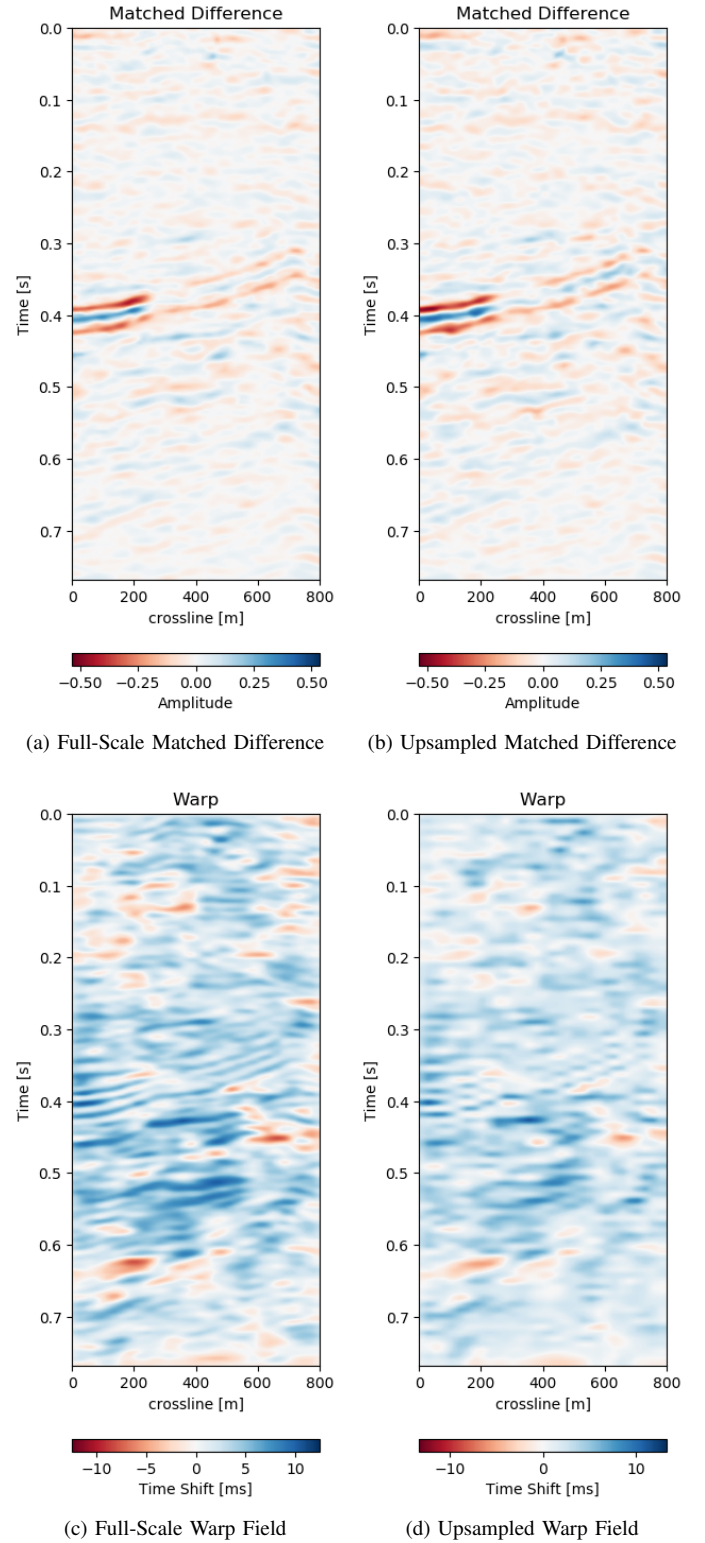


Fig. 6. Warp field with uncertainty that accompanies Figure 9

that the network handles faults well in both training recall and test data, that in theory could violate the diffeomorphic assumption.

We go on to compare a full-scale network to an upsampled network. The full-scale network yields better results and is preferable on seismic data in comparison to the upsampled network presented in the original medical Voxelmorph.

We do expect the network to improve upon training on a



more diverse variety of data sets and seismic responses. While the initial training is time-consuming (25 h), inference is near instantaneous. Moreover, transfer of the trained network to a new data set is possible without training, while accepting some error. Alternatively fine-tuning to new data is possible within few epochs (<1 h).

#### ACKNOWLEDGMENT

The research leading to these results has received funding from the Danish Hydrocarbon Research and Technology Centre under the Advanced Water Flooding program. We thank DTU Compute for access to the GPU Cluster.

#### REFERENCES

- [1] C. MacBeth, M.-D. Mangriotis, and H. Amini, "Post-stack 4d seismic time-shifts: Interpretation and evaluation," *Geophysical Prospecting*, vol. 67, no. 1, pp. 3–31, 2019.
- [2] A. Wirgin, "The inverse crime," *arXiv preprint math-ph/0401050*, 2004.
- [3] P. Hatchell and S. Bourne, "Rocks under strain: Strain-induced time-lapse time shifts are observed for depleting reservoirs," *The Leading Edge*, vol. 24, no. 12, pp. 1222–1225, 2005.
- [4] D. Hale, "Dynamic warping of seismic images," *Geophysics*, vol. 78, no. 2, pp. S105–S115, 2013.
- [5] E. Zabihi Naeini, H. Hoeber, G. Poole, and H. R. Siahkoobi, "Simultaneous multivintage time-shift estimation," *Geophysics*, vol. 74, no. 5, pp. V109–V121, 2009.
- [6] D. Hale, "An efficient method for computing local cross-correlations of multi-dimensional signals," *CWP Report*, vol. 656, 2006.
- [7] J. Rickett, L. Duranti, T. Hudson, B. Regel, and N. Hodgson, "4d time strain and the seismic signature of geomechanical compaction at genesis," *The Leading Edge*, vol. 26, no. 5, pp. 644–647, 2007.
- [8] A. Cherrett, I. Escobar, and H. Hansen, "Fast deterministic geostatistical inversion," in *73rd EAGE Conference and Exhibition incorporating SPE EUROPEC 2011*, 2011.
- [9] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2758–2766.
- [10] A. Ranjan and M. J. Black, "Optical flow estimation using a spatial pyramid network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4161–4170.
- [11] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2462–2470.
- [12] T.-W. Hui, X. Tang, and C. Change Loy, "Liteflownet: A lightweight convolutional neural network for optical flow estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8981–8989.
- [13] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8934–8943.
- [14] D. Gadot and L. Wolf, "Patchbatch: A batch augmented loss for optical flow," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4236–4245.
- [15] F. Güney and A. Geiger, "Deep discrete flow," in *Asian Conference on Computer Vision*. Springer, 2016, pp. 207–224.
- [16] S. Chopra, R. Hadsell, Y. LeCun *et al.*, "Learning a similarity metric discriminatively, with application to face verification," in *CVPR (1)*, 2005, pp. 539–546.
- [17] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "Deepflow: Large displacement optical flow with deep matching," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1385–1392.
- [18] S. Meister, J. Hur, and S. Roth, "Unflow: Unsupervised learning of optical flow with a bidirectional census loss," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [19] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *European conference on computer vision*. Springer, 1994, pp. 151–158.
- [20] F. Stein, "Efficient computation of optical flow using the census transform," in *Joint Pattern Recognition Symposium*. Springer, 2004, pp. 79–86.
- [21] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- [22] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [23] L. Mosser, W. Kimman, J. Dramsch, S. Purves, A. De la Fuente Briceño, and G. Ganssle, "Rapid seismic domain transfer: Seismic velocity inversion and modeling using deep generative neural networks," in *80th EAGE Conference and Exhibition 2018*, 2018.
- [24] A. Bansal, S. Ma, D. Ramanan, and Y. Sheikh, "Recycle-gan: Unsupervised video retargeting," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 119–135.
- [25] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE transactions on medical imaging*, 2019.
- [26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [27] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning for fast probabilistic diffeomorphic registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 729–738.
- [28] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [29] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: <http://tensorflow.org/>
- [30] F. Chollet *et al.*, "Keras," <https://keras.io>, 2015.
- [31] A. V. Dalca, G. Balakrishnan, B. Fischl, P. Golland, J. Guttag, J. E. Iglesias, M. Rakic, M. R. Sabuncu, E. Yu, A. Zhao *et al.*, "Voxelmorph," [voxelmorph.mit.edu](http://voxelmorph.mit.edu), 2018.

**Michael Shell** Biography text here.

PLACE  
PHOTO  
HERE

**John Doe** Biography text here.

**Jane Doe** Biography text here.

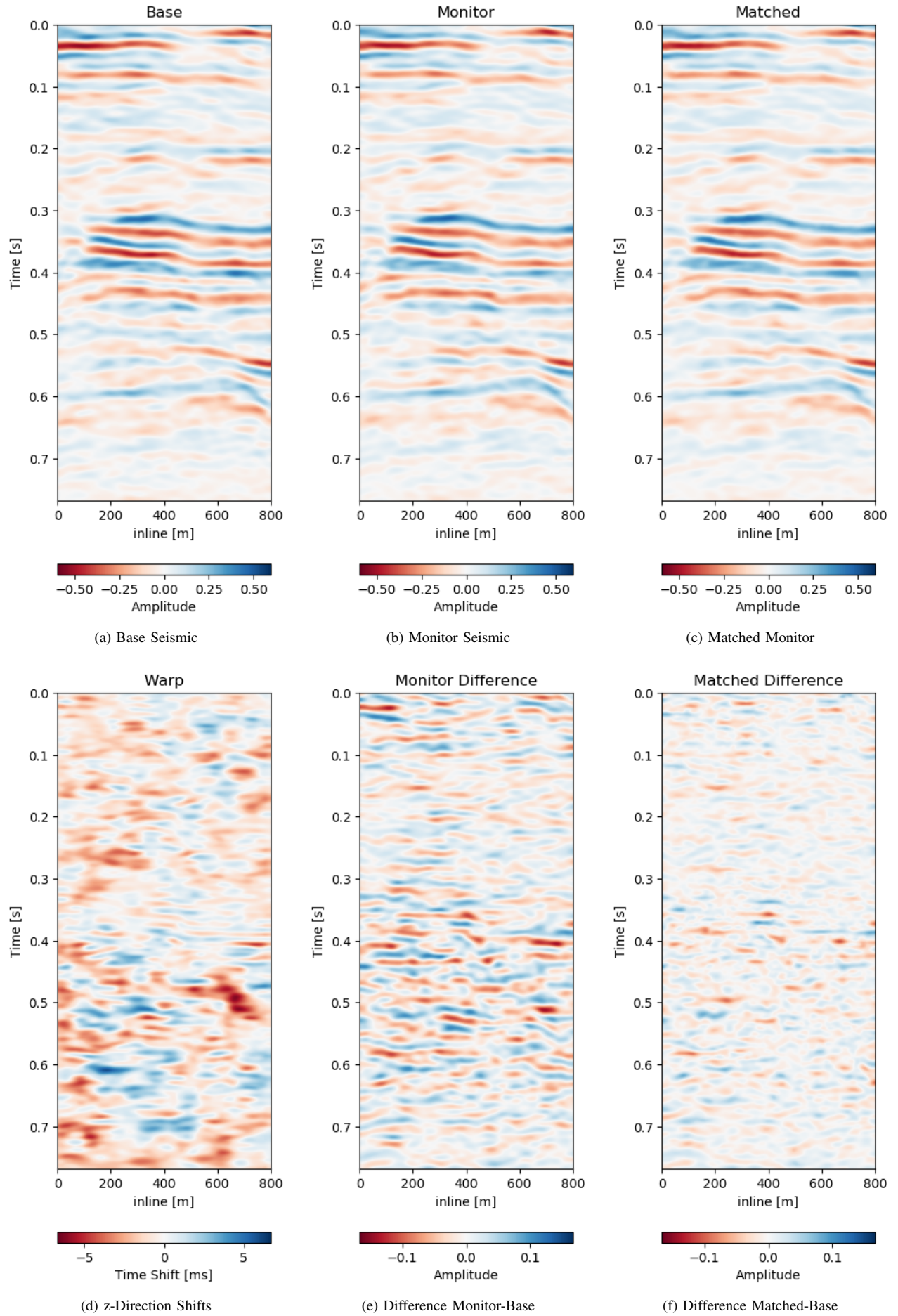


Fig. 7. Generalization of network to same field with different data.

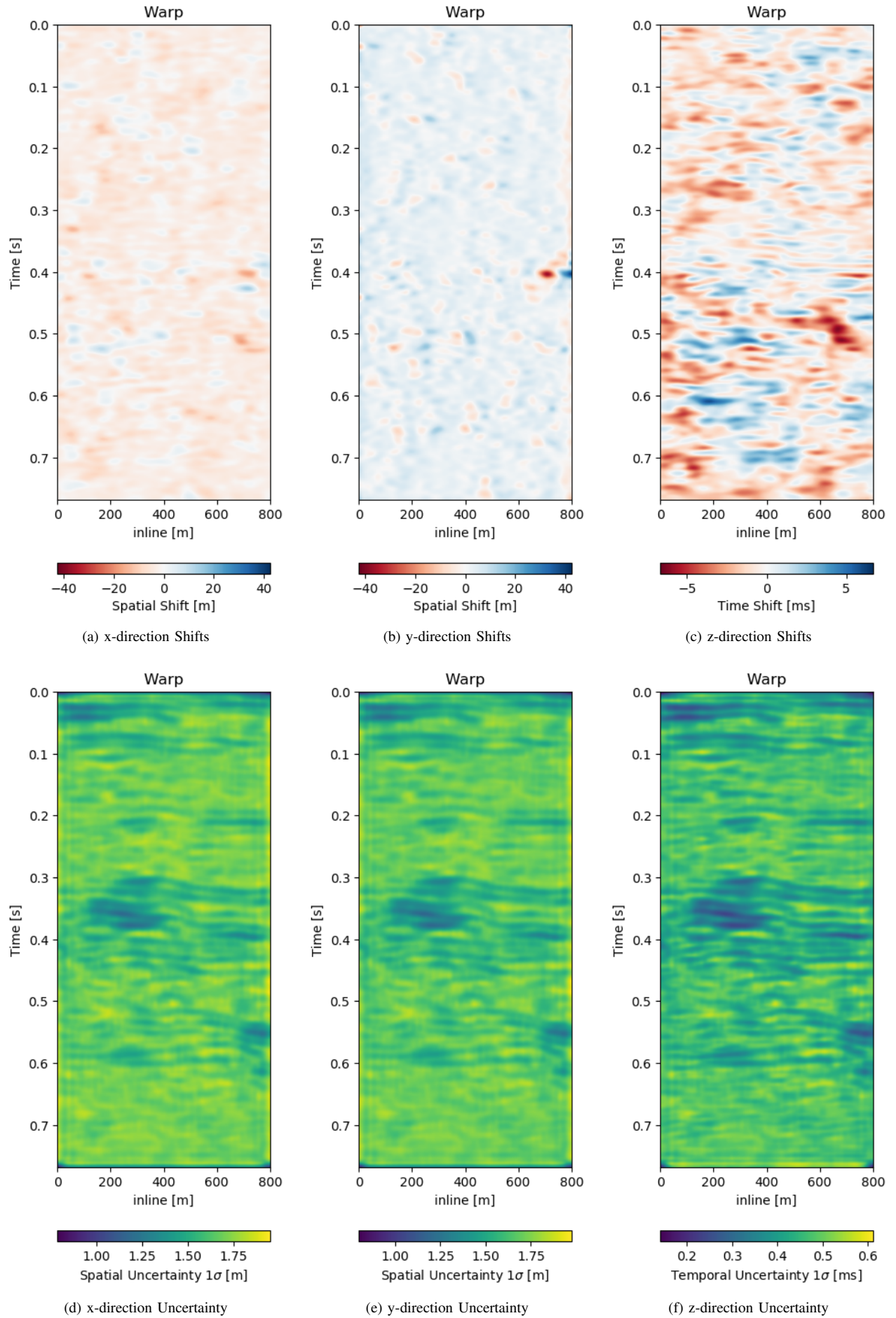


Fig. 8. Warp field with uncertainty that accompanies Figure 7



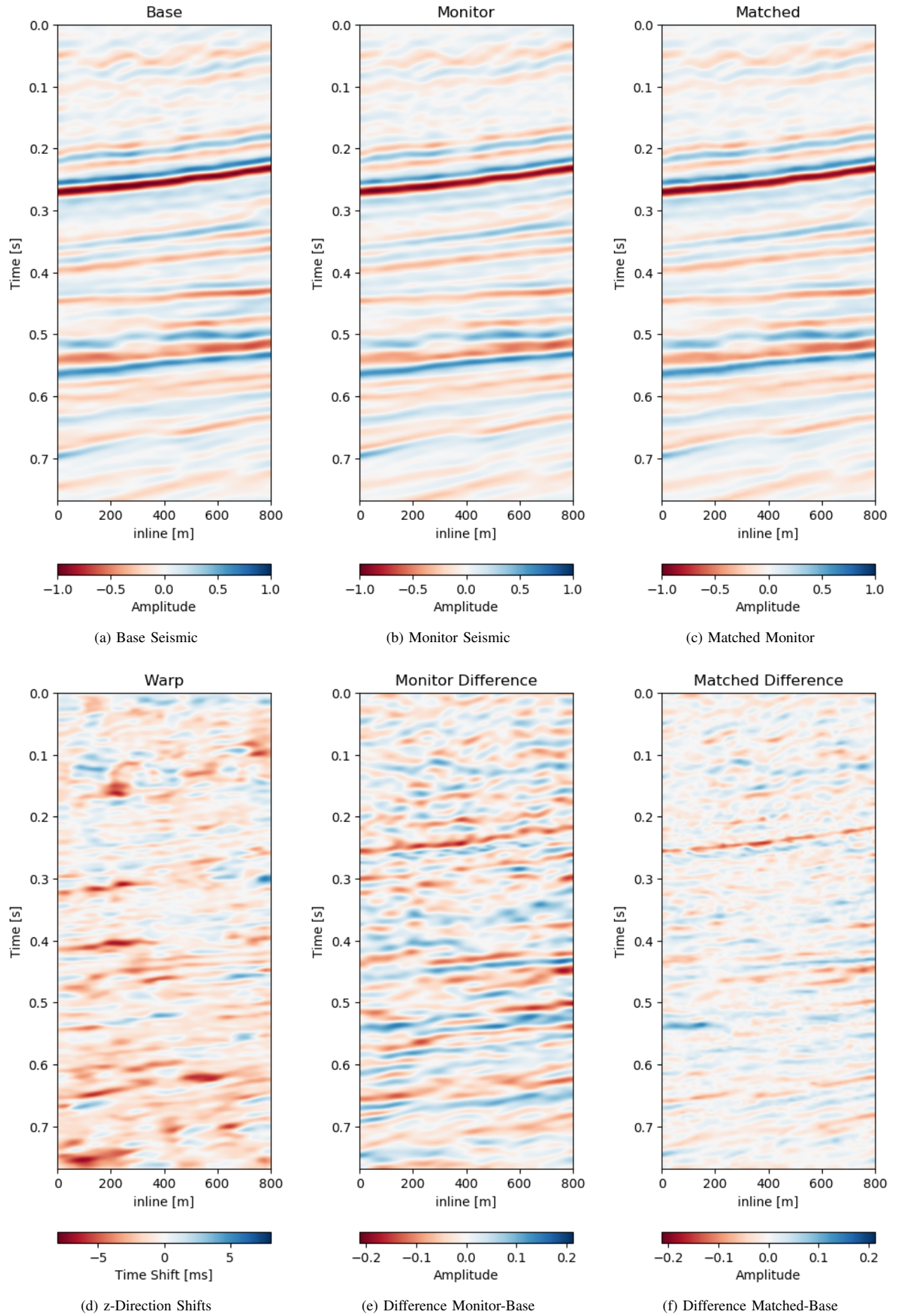


Fig. 9. Generalization of network to a different field.



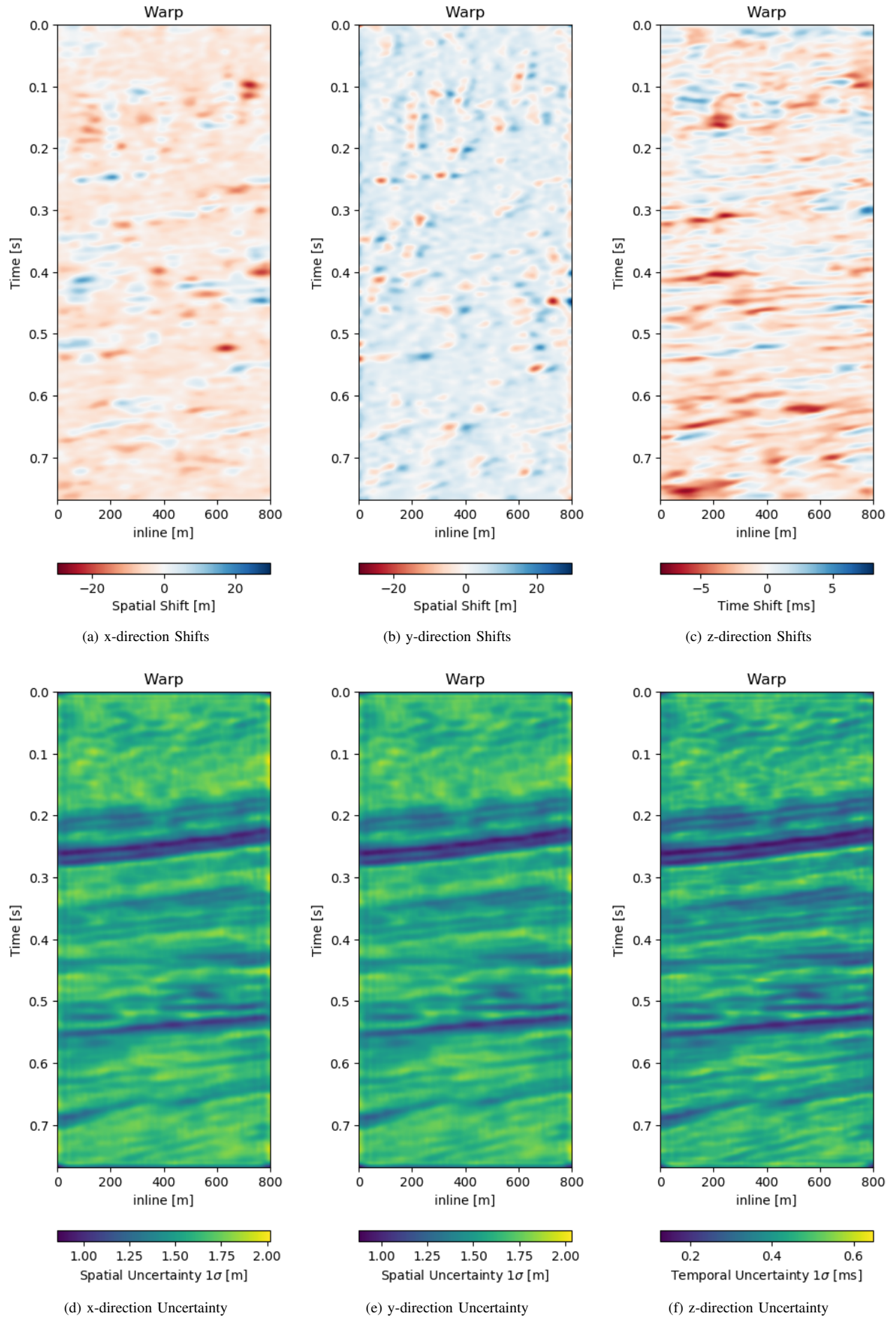


Fig. 10. Warp field with uncertainty that accompanies Figure 9