Global predicted bathymetry using neural networks

Hugh Harper[1] and David T. Sandwell[2]

[1]UC San Diego
[2]UCSD

August 4, 2023

Abstract

Sparse distribution of depth soundings in the ocean make it necessary to infer depth in the gaps using alternate information such as satellite-derived gravity and a mapping from gravity to depth. We design and train a neural network on a collection of 50 million depth soundings to predict bathymetry globally using gravity anomalies. We find the best result is achieved by pre-filtering depth and gravity in accordance with isostatic admittance theory described in previous predicted depth studies. When training the model, if the training and testing split is a random partition at the same resolution as the data, the training and testing sets will not be independent, and model misfit results will be too optimistic. We solve this problem by partitioning the training and testing set with geographic bins. Our final predicted depth model improves on old predicted depth model rms by 16%, from 165 m to 138 m. Among constrained grid cells, 80% of the predicted values are within 128 m of the true value. Improvements to this model will continue with additional depth measurements, but higher resolution predictions, being limited by upward continuation of gravity, shouldn't be attempted with this method.

Global predicted bathymetry using neural networks

Hugh Harper[1], **David T. Sandwell**[1]

[1]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

Key Points:

- We present a new method for global bathymetry prediction using a machine learning algorithm.
- The new predicted depth model improves on the reference model by all error metrics.

Corresponding author: Hugh Harper, huharper@ucsd.edu

Abstract

Sparse distribution of depth soundings in the ocean make it necessary to infer depth in the gaps using alternate information such as satellite-derived gravity and a mapping from gravity to depth. We design and train a neural network on a collection of 50 million depth soundings to predict bathymetry globally using gravity anomalies. We find the best result is achieved by pre-filtering depth and gravity in accordance with isostatic admittance theory described in previous predicted depth studies. When training the model, if the training and testing split is a random partition at the same resolution as the data, the training and testing sets will not be independent, and model misfit results will be too optimistic. We solve this problem by partitioning the training and testing set with geographic bins. Our final predicted depth model improves on old predicted depth model rms by 16%, from 165 m to 138 m. Among constrained grid cells, 80% of the predicted values are within 128 m of the true value. Improvements to this model will continue with additional depth measurements, but higher resolution predictions, being limited by upward continuation of gravity, shouldn't be attempted with this method.

Plain Language Summary

Only a fraction of the seafloor has been mapped by shipboard means. In the unmapped regions of the ocean, we must estimate the depth of the seafloor using information from the earth's gravity field. Typical models of predicted depth determine the linear relationship of gravity and depth in some region, and regional predictions are combined to make global predicted depth maps. Here, we describe a new method for predicting depth from globally using gravity, decades of shipboard depth measurements, and a neural network regression. Ultimately, our model shows a clear improvement over the reference model.

1 Introduction

Less than 25% of the ocean floor has been mapped at 15 arcsecond resolution (GEBCO compilation group, 2023), and at 1 arcminute resolution, this figure is still less than 30%. From efforts such as the Seabed 2030 project (Mayer et al., 2018), coverage in publicly-available compilations has improved in recent years, but the distribution of shipboard depth measurements remains heterogeneous and sparse, providing nearly complete high resolution coverage in some coastal areas but leaving unmapped gaps the size of western US states in remote regions (Figure 1a). While there is no substitute for shipboard surveys to recover high resolution bathymetry, we can make a good guess of the seafloor depth, at a limited resolution, in these gaps using gravity field data derived from satellite altimeters (e.g. Smith & Sandwell, 1994).

Satellite measurements have provided a wealth of information on the gravity field with global coverage, at a resolution of 12 km and accuracy nearing 1 mgal (Sandwell et al., 2021). A new generation of swath altimeters will improve the resolution of the gravity field may improve the accuracy beyond 1 mgal. Since gravity anomaly and depth are correlated within certain wavelength bands (Smith, 1998), we can infer depth from gravity. Within a restricted region, depths may be directly inverted from gravity measurements (e.g. Parker, 1972), but this requires a priori knowledge of crustal density other geologic quantities such as the degree of isostatic compensation (Watts, 2001). There are limitations to this method preventing its application at very large scales. How exactly one should combine sparse depth measurements with global gravity measurements to generate global predicted depths is not trivial.

Smith and Sandwell (1994) developed (and revised in Smith and Sandwell (1997)) an algorithm for this purpose with tremendous success. The procedure uses admittance theory to design filters for bathymetry and gravity to linearize their relationship. The filtering steps remove most assumptions of isostatic compensation. After filtering, they determine the local slope of the bathymetry-gravity relationship on a coarse grid–full details may be found in Smith and Sandwell (1994)–and the predicted bathymetry is the product of the filtered gravity and slope. This paper terms the post-filtering process the "inverse Nettleton," referring to Nettleton (1939) to describe the process of selecting a best-fitting slope to describe the relationship of gravity anomaly and topography. As such, we will refer to this method as the Nettleton method. The quality of the estimation has improved with increasingly precise gravity recovery (number of measurements, orbits, instrument quality, processing techniques) and the greater quantity of shipboard measurements, especially in the so-called gaps. The most recent iteration of this prediction is described in detail by Tozer et al. (2019). While the predicted depth product has been widely adopted, some of the details in the prediction method may not be optimal and leave room for improvement.

There has been recent interest in using modern methods from machine learning to improve upon the prediction of bathymetry. For example (Annan & Wan, 2022) and (Wan et al., 2023) have used neural networks with various architectures to predict absolute depth from gravity and gravity-related quantities (e.g. deflections of the vertical, gravity gradients). An investigation by (Moran et al., 2022) tested many machine learning algorithms in an attempt to predict depth from a set of geophysical and oceanographic features. These models are all limited to a particular study area–training and predictions are restricted to selected regions. The present study is an attempt to update the global predicted depth grid, and our goal is to replace the Nettleton method of depth estimation with a new approach using techniques from machine learning. Specifically, we will train a deep neural network (DNN) to predict depth using a publicly-available collection of depth measurements. We distinguish our method from previous predicted bathymetry studies in a few key ways: we attempt a global prediction; we predict depth in a certain waveband rather than the absolute depth; and we split training and testing data in a unique way.

2 Methods

2.1 Data preparation and feature generation

We begin with the collection of shipboard depth measurements. The collection is based on the collection used in the SRTM15+V2 product (Tozer et al., 2019), and a description of the data sources is found in that study and Becker et al. (2009). Since Tozer et al. (2019), data from 905 cruises (retrieved from NCEI) have been added to the collection. Data have been manually edited to remove erroneous measurements. For our purposes, data provenance is treated equally, and data are not distinguished by cruise ID or instrument type (multibeam or single-beam). Raw shipboard data are reduced by a median filter to 15 arcsecond resolution. These data are combined and blockmedian-reduced to 1 arcminute resolution on a spherical Mercator projected grid, spanning -80.738° to 80.738° latitude. The result is a collection of 52,253,670 records of type [longitude, latitude, depth] (or $[\phi, \theta, d]$).

We can use the coordinates ϕ, θ of any constrained depth record to sample the global gravity anomaly grid (Figure 1c) (Sandwell et al., 2021). Since the constrained depth cells are co-registered with the gravity grid, sampling is trivial, but sampling may also be done via interpolation. The result is records of type $[\phi, \theta, d, g]$. From these records, the target quantity we wish to predict is depth, and the features we may use to train the prediction are ϕ, θ, and g. We could use other geophysical or geographical grids as features, since they can be sampled by longitude and latitude. We will explore such additional features in the discussion.

That longitude is cyclical is not captured by the simple numerical value, so we decompose the longitude, ϕ, to $\sin\left(\frac{\phi\pi}{180}\right)$ and $\cos\left(\frac{\phi\pi}{180}\right)$, or ϕ_s, ϕ_c. With this treatment of

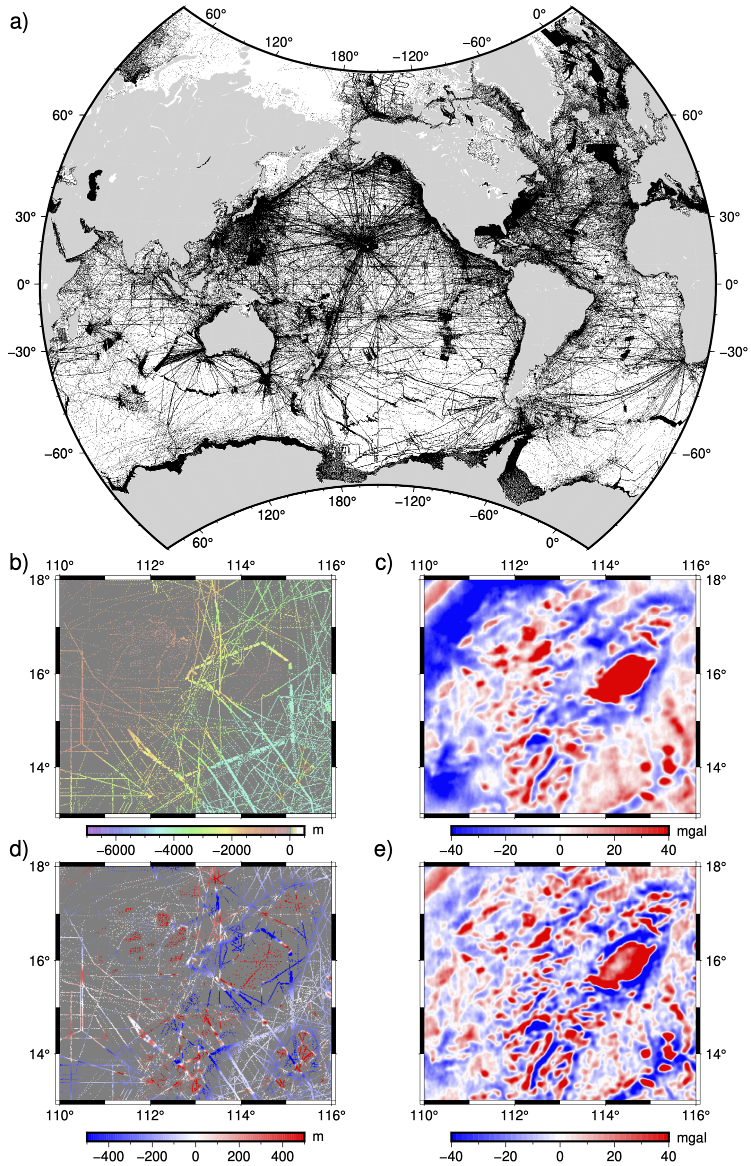


Figure 1. An overview of the datasets in map view. (a) Distribution of shipboard depth measurements in the global oceans based on publicly-available data. (b) Zoomed-in view of depth measurements in the South China Sea, colored by absolute depth. (c) Free air gravity anomaly for the same region as (b). (d) High-pass filtered depths and (e) filtered gravity anomaly–filtering described in the text.

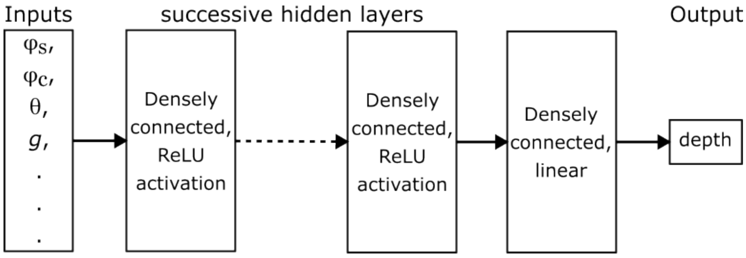


Figure 2. Schematic of the neural network architecture. A feature vector with normalized inputs is transformed by successive hidden layers to predict depth.

longitude, we avoid a discontinuity at the Greenwich Meridian in the predicted depth grid. Following this amendment, our feature vector is $[\phi_s, \phi_c, \theta, g]$.

2.1.1 Filtering depth and gravity

Since gravity and bathymetry are only correlated over certain wavelengths (Smith, 1998), it would be less-than-optimal to try and predict the absolute depth. Here we use the filtering steps established by Smith and Sandwell (1994). The depth measurements are gridded on a 1 arcminute spherical Mercator grid using continuous splines in tension (Smith & Wessel, 1990), and this grid is separated into low- and high-frequency components with a Gaussian filter with 0.5 gain at 160 km (eq. 9, (Smith & Sandwell, 1994)). The low-pass depth grid is saved. The high-pass depth is then low-pass filtered with 0.5 gain at 16 km, resulting in a 160 km - 16 km band-pass depth grid–we will call this h. The constrained points of this depth grid are extracted (Figure 1d). We are losing some high-frequency information this way in order to match the spectral content of the gravity. The gravity anomaly is high-pass filtered in the same way as the depth measurements. Finally, the high-pass filtered gravity anomaly is downward-continued to the low-pass filtered depth using a depth-dependent Wiener filter (eq. 11, (Smith & Sandwell, 1994))– we will call this g^* (Figure 1e). We will examine the effects of omitting this pre-processing step in the discussion.

2.2 Data splitting

We must split our dataset into training, validation, and testing sets. If we are to simply select 20% for testing and validation at random, then we will find that almost any given record in the testing or validation set is within 1 arc minute of a record in the training set. Marks et al. (2010), analyzing predicted depths generated by the Nettleton procedure, showed the prediction error increases with distance from constrained nodes. How strong this effect is depends on the roughness of the seafloor, but it appears the errors become decorrelated beyond a certain distance (15-30 km). In other words, records that are sufficiently close in position are not independent (Figure 3b). In practice, the training loss and validation (and testing) loss will be nearly identical, and we will not have a good idea of when the model is overfitting during training or how the model generalizes to the unmapped gaps.

By splitting the data into longitude, latitude bins and randomly selecting bins for testing and validation, we can reduce the dependence of the datasets (Figure 3c). We use a bin size of 30 arc minutes ($\sim$ 50 km at the equator) to group records, and then randomly select those bins for training, validation, and testing. This bin size could be made larger or smaller, or it could be made to vary based on prior knowledge of seafloor roughness, but it's important not to tune the bin size by model loss performance. The

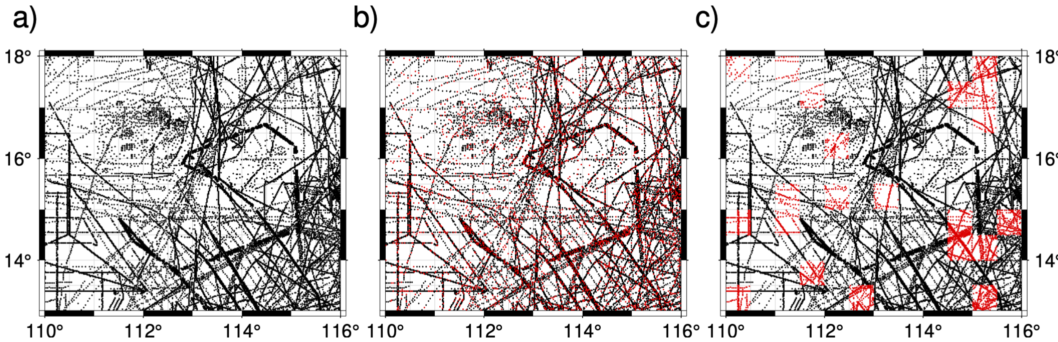


Figure 3. An example of partitioning the data. Same map area shown in Figure 1b-e. (a) The collection of depth measurements. The data must be partitioned into a training, testing, and validation set. (b) Randomly withholding 20% of the data, sampled uniformly. Withheld data are shown in red. Using this partition scheme, almost any withheld point has a nearly identical point in the training set, so the sets are not independent. (c) Sampling the data after binning into groups of 30 arc minutes. Withheld data are shown in red.

training, validation, and testing datasets comprise 31,320,896, 10,460,197, and 10,472,576 records respectively.

2.3 Model architecture and training

We use the TensorFlow software library to design and train the neural network (Abadi et al., 2015). The neural network comprises only successive densely-connected layers using a ReLU activation function (Figure 2). Input features are normalized by the mean and variance of their distribution in the training dataset. We use eight successive dense layers with 256 neurons per layer, and a final linear output layer. Model architecture can be tweaked ad nauseum, so we cannot claim this is a strictly optimal configuration, but this particular arrangement was selected because we found it performed as well as a wider but shallower model (e.g., 4 layers of 1024 neurons each) while using far fewer parameters (4e5 vs. 3e6).

We choose mean squared error (MSE) as the loss function. Each dense layer is regularized with L2 regularization ($\lambda = 0.01$). We use the Adam optimizer (Kingma & Ba, 2017) with a learning rate of 0.001. Model training proceeds until the validation loss is no longer decreasing.

2.4 Inference: generating the global predicted grid

The end goal of this model is a global grid of predicted depths. After the model is trained, predictions of h are generated on a 1 arcminute spherical mercator grid (onshore values are masked). The long wavelength depth, saved from the filtering step, are added to the predicted depth, h, to give absolute depth, d. Finally, for distribution, the predicted depth grid is "polished" with the constrained depth measurements, but this step is omitted in the following discussion and analysis.

3 Results

3.1 Base model

Using the feature vector $[\phi_s, \phi_c, \theta, g^*]$ to predict h, we achieve a training RMSE of 85 m, validation RMSE of 108 m, and testing RMSE of 109 m. Loss values are useful for comparing one trained model against another, but they are imperfect when comparing to the Nettleton method. Since no "testing" data is withheld from the Nettleton method (i.e., the prediction is tuned on all available data), we must caution the comparison of RMSE between the two methods. With that in mind, the RMSE of the Nettleton prediction and h is about 143 m. We will look more carefully at model misfit in the discussion.

3.1.1 Modeling without filtering depths and gravity

For comparison, a model trained without filtering depths and gravity anomaly performs much worse and offers no improvement over the Nettleton method. For this model, we achieve a training RMSE of 140 m, validation RMSE of 173 m, and testing RMSE of 175 m. This poor performance may be because the distribution of depth is more variable with location. For example, where the regional depth is 6000 m, the mean depth will be near 6000 m, and similar for a regional depth of 1000 m. By high-pass filtering the depth, the overall variance of the data are reduced. Omitting the low-pass filter at 16 km also contributes to the worse performance. The Nettleton method RMSE above is for the band-pass filtered data, h. If the short wavelengths are included, the Nettleton RMSE is about 165 m.

Alternatively, we can high-pass filter depth and gravity but omit the low-pass filter at 16 km–in fact, we may desire to do this so we don't lose short-wavelength details. For this model, we achieve a training RMSE of 124 m, validation RMSE of 141 m, and testing RMSE of 142 m. We can't compare these directly to the base model since the target quantities are different. Instead we can evaluate the RMSE of the base model prediction and the high-pass depth. In this case, testing RMSE values are nearly identical for the two models, suggesting the trained models are similar and the greater loss reflects the greater variance of the target data.

3.2 Added features

It would seem that adding features from other global grids would be an easy way to decrease model loss and improve performance. For example, the spreading rate at the time crust is created is known to affect the roughness of bathymetry (Small & Sandwell, 1994). Crustal age and sediment cover will also influence the correlation of gravity and bathymetry (Smith & Sandwell, 1994). We use ϕ, θ to sample grids of crustal age, paleo-spreading rate (Seton et al., 2020), and sediment thickness (Whittaker et al., 2013), add these to the feature vector, and train a new model. In practice, there are problems with using these features.

Firstly, these grids have many regions of missing data, and the missing values must be handled somehow. Since a key purpose of this model is prediction on a global grid, grid cells with missing values can't simply be thrown out. We tested different schemes to replacing missing values: replacement with the mean feature value; replacement with mean feature value plus an additional boolean feature indicating replacement; and filling missing values with nearest neighbor interpolation of the feature grid. In our attempts to use crustal age, paleo-spreading rate, and sediment thickness as features, validation and testing loss are not improved (nor are they improved by any one such feature), and in fact model loss is worse with the additional features. In addition, at inference time, sharp discontinuities in the feature grids get mapped to the prediction grid creating unwanted artifacts.

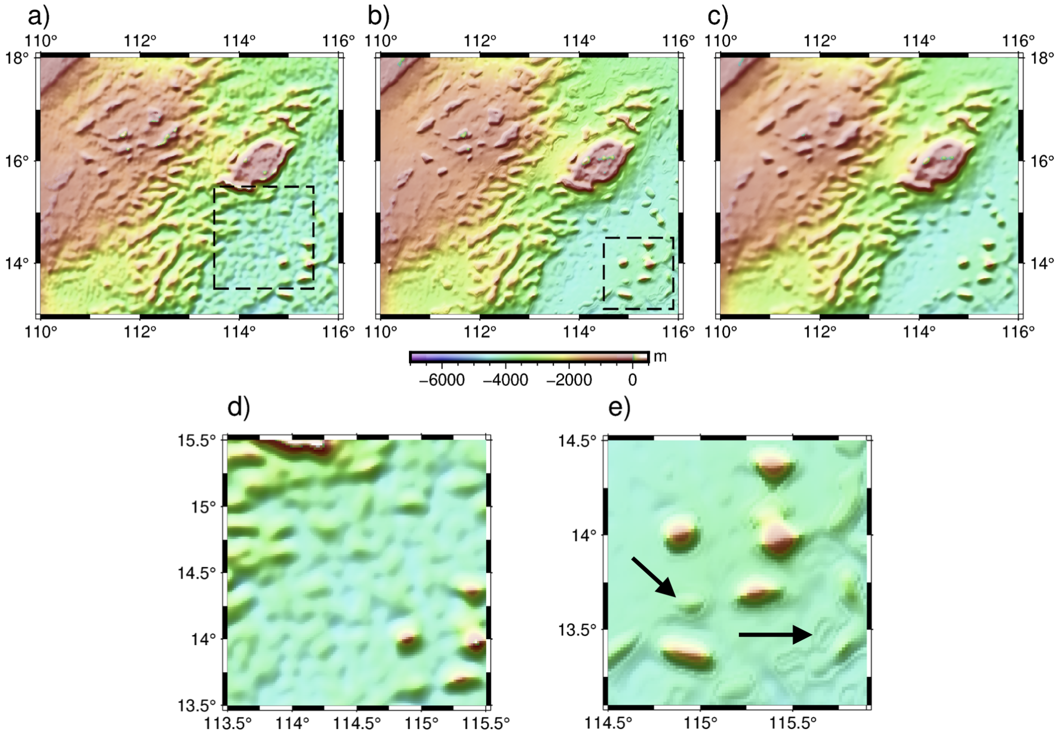


Figure 4. Model-predicted depths with long-wavelength depth added. The same map area shown in Figure 1b-e. (a) Nettleton method; (b) Raw DNN-predicted grid; (c) DNN-predicted grid, low-pass filtered at 12 km. (d) Example of the "orange peel" texture in the Nettleton prediction, boxed region in (a). (e) Example of "hallucinations" in DNN prediction, boxed region in (b).

Other gravity-related quantities such as deflection of the vertical and vertical gravity gradient (VGG) may be good features to use, but they may only be redundant. We found that adding VGG as a feature (Sandwell et al., 2021) improves model training RMSE only slightly, and it slightly worsens validation and testing RMSE. This model has a training RMSE of 85 m, a validation RMSE of 109 m, and a testing RMSE of 110 m. Since there is no improvement on the base model, we will consider that the preferred model and refer to it as simply the "DNN" model in the following discussion.

4 Discussion

4.1 Generating a predicted depth grid

After training, we generate model predicted depths on a global 1 min mercator grid. One problem that results is short wavelength artifacts or "hallucinations" (Figure 4e). These hallucinations typically occur with wavelengths shorter than the 16 km wavelength filter that was applied to the original gravity and bathymetry, so they must be a product of the DNN training. We can reduce these with regularization during training, but not completely nor in a deterministic way. For the distributed predicted depth grid, we apply a low pass filter with 0.5 gain at 16 km to remove these hallucinations. This post-inference filtering method does not weaken model results. In fact, error metrics are very slightly improved, and the prediction RMSE of h after filtering is 107 m for the testing dataset. We use the predictions on the filtered grid in the following discussion.

4.2 Comparison to Nettleton model

While not quantitative, it's important to visually inspect the DNN predicted depth grid and make comparisons to the Nettleton grid (Figure 4). The most obvious qualitative difference is in continental margin areas or areas of relatively smooth seafloor. In these areas, the Nettleton predicted bathymetry has a rough "orange peel" texture (Figure 4d), an artifact of downward continuation of noisy gravity data. This type of seafloor is smoother in the DNN prediction.

Using all available depth measurements (not partitioned for training/testing), we compare the error distribution of the Nettleton method and the DNN method. These results are shown in Figure 5. For the Nettleton method, the predicted depth is within 68 m for 50% of points and within 168 m for 80% of points. For the DNN method, these percentiles are 45 m and 128 m, respectively. Additionally, the mean error has been reduced from 13 m for the Nettleton to 3 m for the DNN, indicating a less-biased estimate. Figure 5b shows the distribution of absolute model error in the southern oceans. Overall, the spatial patterns of misfit are similar for the two models. At this scale, the noticeable differences are found nearer to land–e.g., the West Antarctic Peninsula, Chile, Australia–where the DNN model shows clear improvements over the Nettleton.

4.2.1 BODC data

Because the Nettleton prediction is tuned using all available data, we don't have a concrete idea of how well it generalizes to unseen areas of seafloor. It is useful to reserve a dataset that is not used in either model's construction. To compare the performance of the Nettleton and the DNN model, we used a collection of depth measurements from 279 cruises from the British Oceanographic Data Centre (BODC) that are not yet incorporated into the prediction model. The raw data are decimated to 15 sec, and measurements that overlap with data already in the prediction dataset are removed. We did not thoroughly inspect the BODC data for erroneous measurements, so measurements that differ from either predicted grid by more than 2000 m are removed. In total, there are 6,242,414 points. The Nettleton prediction has an RMSE of 150 m for the BODC dataset. The final DNN model has an RMSE of 104 m.

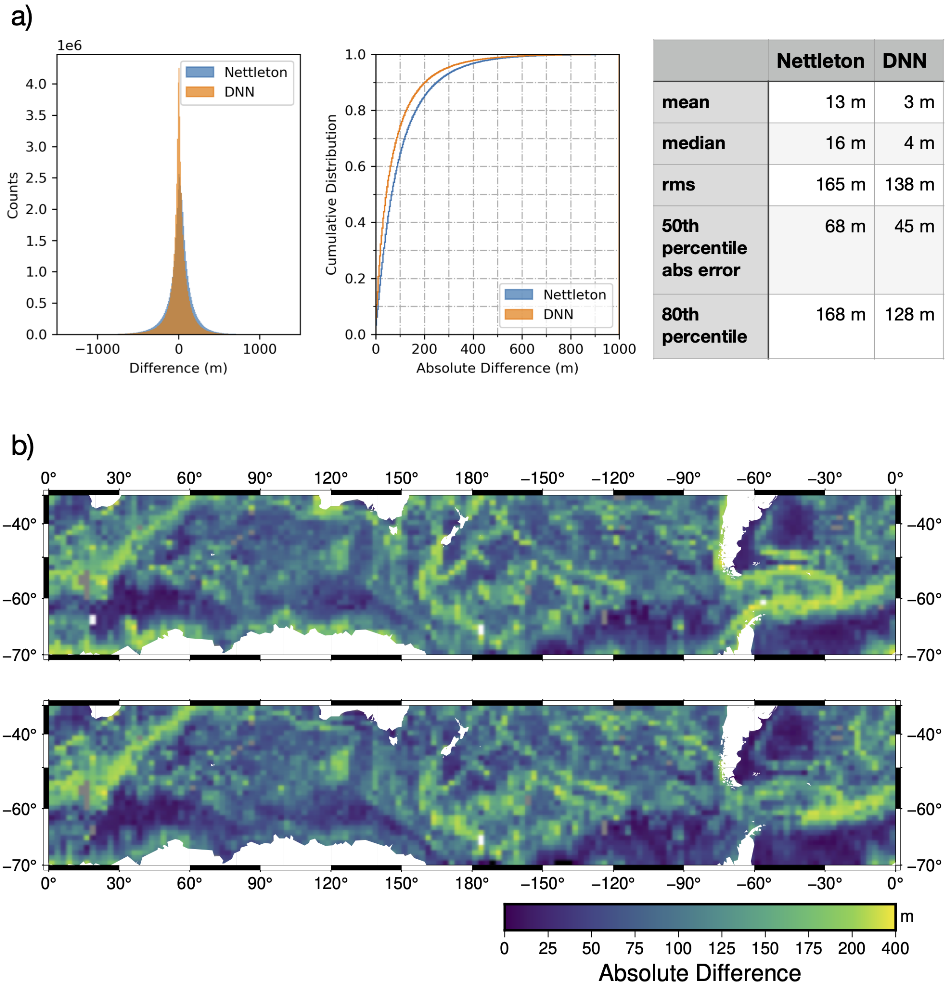
If we restrict the analysis spatially to the highest concentration of measurements (80% of the data are around the British Isles), the Nettleton RMSE is 73 m and the DNN model RMSE is 62 m–much lower than testing RMSE (Figure 6). This almost certainly reflects the proximity of these data to those in our training set. However, we see from the error distribution that the slight bias in the Nettleton prediction is not present in the DNN prediction.

4.3 Potential for improvements

Our model is a simple implementation of a neural network to predict depth globally, and we have shown its clear improvement over the Nettleton method. Yet, there are many possible directions for improvement depending on one's objectives. Expansion of the training dataset, modifications of model architecture, or a multi-regional approach to the problem all offer potential to improve on our model.

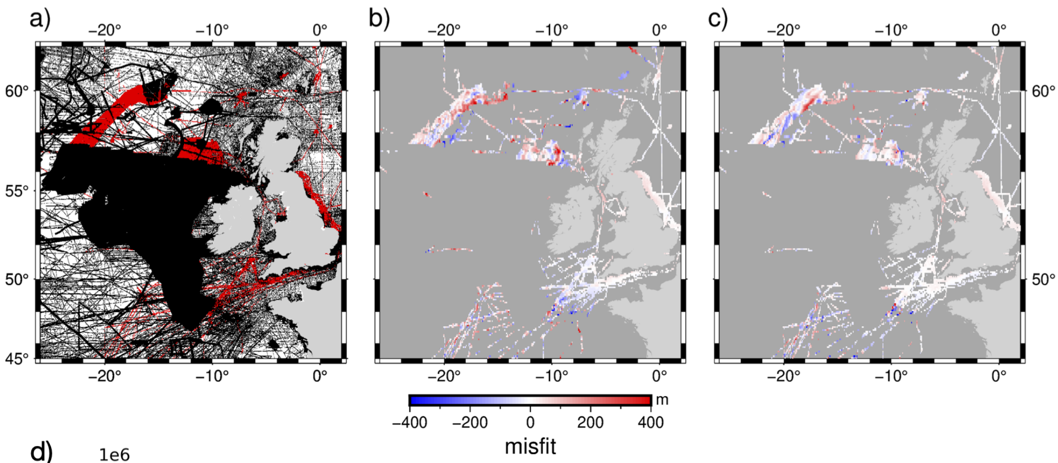
If coverage of publicly available bathymetry compilations continues to improve as it has in recent years–and it likely will (Mayer et al., 2018)–model predictions will clearly improve (this would be true of any model). Low-resolution data in remote regions, which can be collected by autonomous vehicles, will likely offer the greatest benefit in our model approach.

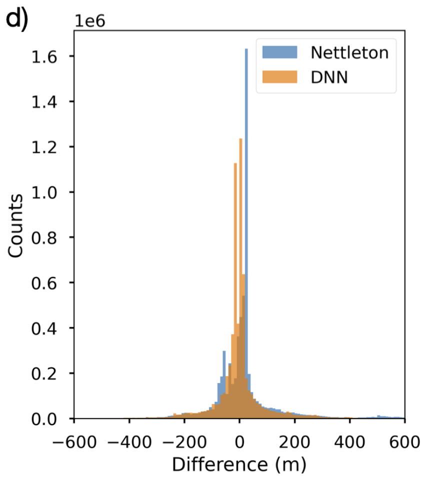
We haven't made use of high-resolution multibeam data in our model, and we do not aim to predict features at such resolution. Upward continuation of gravity anoma-



	Nettleton	DNN
mean	13 m	3 m
median	16 m	4 m
rms	165 m	138 m
50th percentile abs error	68 m	45 m
80th percentile	168 m	128 m

Figure 5. Comparison of Nettleton and DNN models by prediction error. (a) Distribution of prediction error for all 1 arc minute data (N=52,253,670). (b) Average absolute difference (prediction - measurement) for Nettleton (upper) and DNN (lower) models.





	All		British Isles	
	Nettleton	DNN	Nettleton	DNN
mean	22 m	-3 m	12 m	-1 m
median	19 m	-3 m	20 m	-1 m
rms	150 m	104 m	73 m	62 m
50th percentile abs error	28 m	16 m	25 m	14 m
80th percentile	94 m	71 m	60 m	40 m

Figure 6. Nettleton and DNN model predictions for withheld BODC data. (a) Depth measurements in the full dataset shown in black, measurements from the disjoint BODC dataset shown in red. Misfit of BODC data for (b) Nettleton predicted depth and (c) neural network predicted depth. (d) Distribution of BODC data misfit for Nettleton predicted depth and neural network predicted depth.

lies limits the resolution of gravity from satellite altimetry to a length scale of about π times the regional depth (e.g. Smith & Sandwell, 2004), so it is not possible to realistically predict depth from only gravity (and its derivatives) at such scales. An approach using convolutional neural networks, as demonstrated by Annan and Wan (2022), may successfully learn from higher resolution bathymetry in regional settings.

A prediction model trained on regional data will, everything else equal, perform better in that region than a model trained on global data. Moran et al. (2022) identified regions where various learning algorithms might preferentially excel. This suggests a global model should alternatively be constructed from a suite of regional models. A particular case where such a multi-regional model would excel is in predicting higher resolution depth in areas where that is realistic. Susa (2022) showed such an approach to predicting depth in near-coastal regions. In this setting, altimetric ranging and gravity accuracy suffer from land contamination (Raney & Phalippou, 2011), and visible spectra may be correlated with bathymetry, making this an ideal case for alternative depth prediction.

5 Conclusions

1. Using a large collection of depth measurements and satellite-derived gravity anomalies, we trained a deep neural network to predict seafloor depth.
2. We find that applying filters (described by Smith and Sandwell (1994)) to bathymetry and gravity before training is necessary for a good result, and conforms the data more closely with the assumption of identical distributions.
3. When dealing with sparse heterogeneous sampling, the training-testing split must be treated carefully. If the training and testing split is a random partition at the same resolution as the data, the training and testing sets are not independent, and model misfit results will be too optimistic.
4. Adding features sampled from geologic grids–crustal age, paleo-spreading rate, sediment thickness–do not improve model results.
5. Our preferred DNN-predicted model improves on the results of the Nettleton procedure, lowering the prediction RMSE from 165 m to 138 m.
6. While improvements will be made with additional depth measurement data, higher resolution predictions are limited by the upward continuation of gravity, so likely shouldn't be attempted with this method.

Open Research Section

Jupyter notebooks and data files to reproduce the predicted depth model, as well as the final predicted depth model used in the analysis, are available at `https://doi.org/10.5281/zenodo.8029925` (Harper & Sandwell, 2023).

Acknowledgments
We thank Ross Parnell-Turner, Dave May, Jose Restrepo, and Jeff Gee for their suggestions and comments on a preliminary version of the manuscript. This work was supported by the NASA SWOT program (80NSSC20K1138), a NASA FINNEST fellowship (80NSSC20K1616), and the Office of Naval Research (N00014-17-1-2866).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Retrieved from `https://www.tensorflow.org/` (Software available from tensorflow.org)

Annan, R. F., & Wan, X. (2022, 10). Recovering bathymetry of the gulf of guinea using altimetry-derived gravity field products combined via convolutional neural network. *Surveys in Geophysics*, *43*, 1541-1561. doi: 10.1007/s10712-022-09720-5

Becker, J. J., Sandwell, D. T., Smith, W. H., Braud, J., Binder, B., Depner, J., … Weatherall, P. (2009, 10). Global bathymetry and elevation data at 30 arc seconds resolution: Srtm30_plus. *Marine Geodesy*, *32*, 355-371. doi: 10.1080/01490410903297766

GEBCO compilation group. (2023). *Gebco 2023 grid.* doi: 10.5285/f98b053b-0cbc-6c23-e053-6c86abc0af7b

Harper, H., & Sandwell, D. (2023, June). *Global predicted bathymetry using neural networks.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.8029925` doi: 10.5281/zenodo.8029925

Kingma, D. P., & Ba, J. (2017). *Adam: A method for stochastic optimization.*

Marks, K. M., Smith, W. H. F., & Sandwell, D. T. (2010). Evolution of errors in the altimetric bathymetry model used by google earth and gebco. *Marine Geophysical Research*, *31*, 223-238. doi: 10.1007/s11001-010-9102-0

Mayer, L., Jakobsson, M., Allen, G., Dorschel, B., Falconer, R., Ferrini, V., … Weatherall, P. (2018). The nippon foundation-gebco seabed 2030 project: The quest to see the world's oceans completely mapped by 2030. *Geosciences (Switzerland)*, *8*. doi: 10.3390/geosciences8020063

Moran, N., Stringer, B., Lin, B., & Hoque, M. T. (2022). Machine learning model selection for predicting bathymetry. *Deep-Sea Research Part I: Oceanographic Research Papers*, *185*, 103788. doi: 10.1016/j.dsr.2022.103788

Nettleton, L. L. (1939). Determination of density for reduction of gravimeter observations. *Geophysics*, *4*.

Parker, R. (1972). The rapid calculation of potential anomalies. *Geophysical Journal of the Royal Astronomical Society*, *31*, 447-455. doi: 10.1111/j.1365-246X.1973.tb06513.x

Raney, R. K., & Phalippou, L. (2011). The future of coastal altimetry. In S. Vignudelli, A. G. Kostianoy, P. Cipollini, & J. Benveniste (Eds.), *Coastal altimetry* (pp. 535–560). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-12796-0_20

Sandwell, D. T., Harper, H., Tozer, B., & Smith, W. H. F. (2021). Gravity field recovery from geodetic altimeter missions. *Advances in Space Research*, *68*(2), 1059–1072.

Seton, M., Müller, R. D., Zahirovic, S., Williams, S., Wright, N. M., Cannon, J., … McGirr, R. (2020). A global data set of present-day oceanic crustal age and seafloor spreading parameters. *Geochemistry, Geophysics, Geosystems*, *21*, 1-15. doi: 10.1029/2020GC009214

Small, C., & Sandwell, D. T. (1994). Imaging mid-ocean ridge transitions with satellite gravity. *Geology*, *22*, 123-126.

Smith, W. H. F. (1998). Seafloor tectonic fabric from satellite altimetry. *Annual Review of Earth and Planetary Sciences*, *26*, 697-738.

Smith, W. H. F., & Sandwell, D. T. (1994). Bathymetric prediction from dense satellite altimetry and sparse shipboard bathymetry. *Journal of Geophysical Research*, *99*. doi: 10.1029/94jb00988

Smith, W. H. F., & Sandwell, D. T. (1997). Global sea floor topography from satellite altimetry and ship depth soundings. *Science*, *277*, 1956-1962. doi: 10.1126/science.277.5334.1956

Smith, W. H. F., & Sandwell, D. T. (2004). Conventional bathymetry, bathymetry from space, and geodetic altimetry. *Oceanography*, *17*, 8-23. doi: 10.5670/oceanog.2004.63

Smith, W. H. F., & Wessel, P. (1990). Gridding with continuous curvature splines in tension. *Geophysics*, *55*(3), 293–305.

Susa, T. (2022). Satellite derived bathymetry with sentinel-2 imagery: Comparing traditional techniques with advanced methods and machine learning ensemble models. *Marine Geodesy*, *0*, 1-20. doi: 10.1080/01490419.2022.2064572

Tozer, B., Sandwell, D. T., Smith, W. H. F., Olson, C., Beale, J. R., & Wessel, P. (2019). Global bathymetry and topography at 15 arc sec: Srtm15+. *Earth and Space Science*, *6*, 1847-1864. doi: 10.1029/2019EA000658

Wan, X., Annan, R. F., & Ziggah, Y. Y. (2023, 2). Altimetry-derived gravity gradients using spectral method and their performance in bathymetry inversion using back-propagation neural network. *Journal of Geophysical Research: Solid Earth*, *128*. doi: 10.1029/2022JB025785

Watts, A. B. (2001). *Isostasy and flexure of the lithosphere.* Cambridge University Press.

Whittaker, J. M., Goncharov, A., Williams, S. E., Müller, R. D., & Leitchenkov, G. (2013). Global sediment thickness data set updated for the australian-antarctic southern ocean. *Geochemistry, Geophysics, Geosystems*, *14*(8), 3297-3305. doi: https://doi.org/10.1002/ggge.20181

Global predicted bathymetry using neural networks

Hugh Harper[1], **David T. Sandwell**[1]

[1]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

Key Points:

- We present a new method for global bathymetry prediction using a machine learning algorithm.
- The new predicted depth model improves on the reference model by all error metrics.

Corresponding author: Hugh Harper, `huharper@ucsd.edu`

Abstract

Sparse distribution of depth soundings in the ocean make it necessary to infer depth in the gaps using alternate information such as satellite-derived gravity and a mapping from gravity to depth. We design and train a neural network on a collection of 50 million depth soundings to predict bathymetry globally using gravity anomalies. We find the best result is achieved by pre-filtering depth and gravity in accordance with isostatic admittance theory described in previous predicted depth studies. When training the model, if the training and testing split is a random partition at the same resolution as the data, the training and testing sets will not be independent, and model misfit results will be too optimistic. We solve this problem by partitioning the training and testing set with geographic bins. Our final predicted depth model improves on old predicted depth model rms by 16%, from 165 m to 138 m. Among constrained grid cells, 80% of the predicted values are within 128 m of the true value. Improvements to this model will continue with additional depth measurements, but higher resolution predictions, being limited by upward continuation of gravity, shouldn't be attempted with this method.

Plain Language Summary

Only a fraction of the seafloor has been mapped by shipboard means. In the unmapped regions of the ocean, we must estimate the depth of the seafloor using information from the earth's gravity field. Typical models of predicted depth determine the linear relationship of gravity and depth in some region, and regional predictions are combined to make global predicted depth maps. Here, we describe a new method for predicting depth from globally using gravity, decades of shipboard depth measurements, and a neural network regression. Ultimately, our model shows a clear improvement over the reference model.

1 Introduction

Less than 25% of the ocean floor has been mapped at 15 arcsecond resolution (GEBCO compilation group, 2023), and at 1 arcminute resolution, this figure is still less than 30%. From efforts such as the Seabed 2030 project (Mayer et al., 2018), coverage in publicly-available compilations has improved in recent years, but the distribution of shipboard depth measurements remains heterogeneous and sparse, providing nearly complete high resolution coverage in some coastal areas but leaving unmapped gaps the size of western US states in remote regions (Figure 1a). While there is no substitute for shipboard surveys to recover high resolution bathymetry, we can make a good guess of the seafloor depth, at a limited resolution, in these gaps using gravity field data derived from satellite altimeters (e.g. Smith & Sandwell, 1994).

Satellite measurements have provided a wealth of information on the gravity field with global coverage, at a resolution of 12 km and accuracy nearing 1 mgal (Sandwell et al., 2021). A new generation of swath altimeters will improve the resolution of the gravity field may improve the accuracy beyond 1 mgal. Since gravity anomaly and depth are correlated within certain wavelength bands (Smith, 1998), we can infer depth from gravity. Within a restricted region, depths may be directly inverted from gravity measurements (e.g. Parker, 1972), but this requires a priori knowledge of crustal density other geologic quantities such as the degree of isostatic compensation (Watts, 2001). There are limitations to this method preventing its application at very large scales. How exactly one should combine sparse depth measurements with global gravity measurements to generate global predicted depths is not trivial.

Smith and Sandwell (1994) developed (and revised in Smith and Sandwell (1997)) an algorithm for this purpose with tremendous success. The procedure uses admittance theory to design filters for bathymetry and gravity to linearize their relationship. The filtering steps remove most assumptions of isostatic compensation. After filtering, they determine the local slope of the bathymetry-gravity relationship on a coarse grid–full details may be found in Smith and Sandwell (1994)–and the predicted bathymetry is the product of the filtered gravity and slope. This paper terms the post-filtering process the "inverse Nettleton," referring to Nettleton (1939) to describe the process of selecting a best-fitting slope to describe the relationship of gravity anomaly and topography. As such, we will refer to this method as the Nettleton method. The quality of the estimation has improved with increasingly precise gravity recovery (number of measurements, orbits, instrument quality, processing techniques) and the greater quantity of shipboard measurements, especially in the so-called gaps. The most recent iteration of this prediction is described in detail by Tozer et al. (2019). While the predicted depth product has been widely adopted, some of the details in the prediction method may not be optimal and leave room for improvement.

There has been recent interest in using modern methods from machine learning to improve upon the prediction of bathymetry. For example (Annan & Wan, 2022) and (Wan et al., 2023) have used neural networks with various architectures to predict absolute depth from gravity and gravity-related quantities (e.g. deflections of the vertical, gravity gradients). An investigation by (Moran et al., 2022) tested many machine learning algorithms in an attempt to predict depth from a set of geophysical and oceanographic features. These models are all limited to a particular study area–training and predictions are restricted to selected regions. The present study is an attempt to update the global predicted depth grid, and our goal is to replace the Nettleton method of depth estimation with a new approach using techniques from machine learning. Specifically, we will train a deep neural network (DNN) to predict depth using a publicly-available collection of depth measurements. We distinguish our method from previous predicted bathymetry studies in a few key ways: we attempt a global prediction; we predict depth in a certain waveband rather than the absolute depth; and we split training and testing data in a unique way.

2 Methods

2.1 Data preparation and feature generation

We begin with the collection of shipboard depth measurements. The collection is based on the collection used in the SRTM15+V2 product (Tozer et al., 2019), and a description of the data sources is found in that study and Becker et al. (2009). Since Tozer et al. (2019), data from 905 cruises (retrieved from NCEI) have been added to the collection. Data have been manually edited to remove erroneous measurements. For our purposes, data provenance is treated equally, and data are not distinguished by cruise ID or instrument type (multibeam or single-beam). Raw shipboard data are reduced by a median filter to 15 arcsecond resolution. These data are combined and blockmedian-reduced to 1 arcminute resolution on a spherical Mercator projected grid, spanning -80.738° to 80.738° latitude. The result is a collection of 52,253,670 records of type [longitude, latitude, depth] (or $[\phi, \theta, d]$).

We can use the coordinates ϕ, θ of any constrained depth record to sample the global gravity anomaly grid (Figure 1c) (Sandwell et al., 2021). Since the constrained depth cells are co-registered with the gravity grid, sampling is trivial, but sampling may also be done via interpolation. The result is records of type $[\phi, \theta, d, g]$. From these records, the target quantity we wish to predict is depth, and the features we may use to train the prediction are ϕ, θ, and g. We could use other geophysical or geographical grids as features, since they can be sampled by longitude and latitude. We will explore such additional features in the discussion.

That longitude is cyclical is not captured by the simple numerical value, so we decompose the longitude, ϕ, to $\sin\left(\frac{\phi\pi}{180}\right)$ and $\cos\left(\frac{\phi\pi}{180}\right)$, or ϕ_s, ϕ_c. With this treatment of

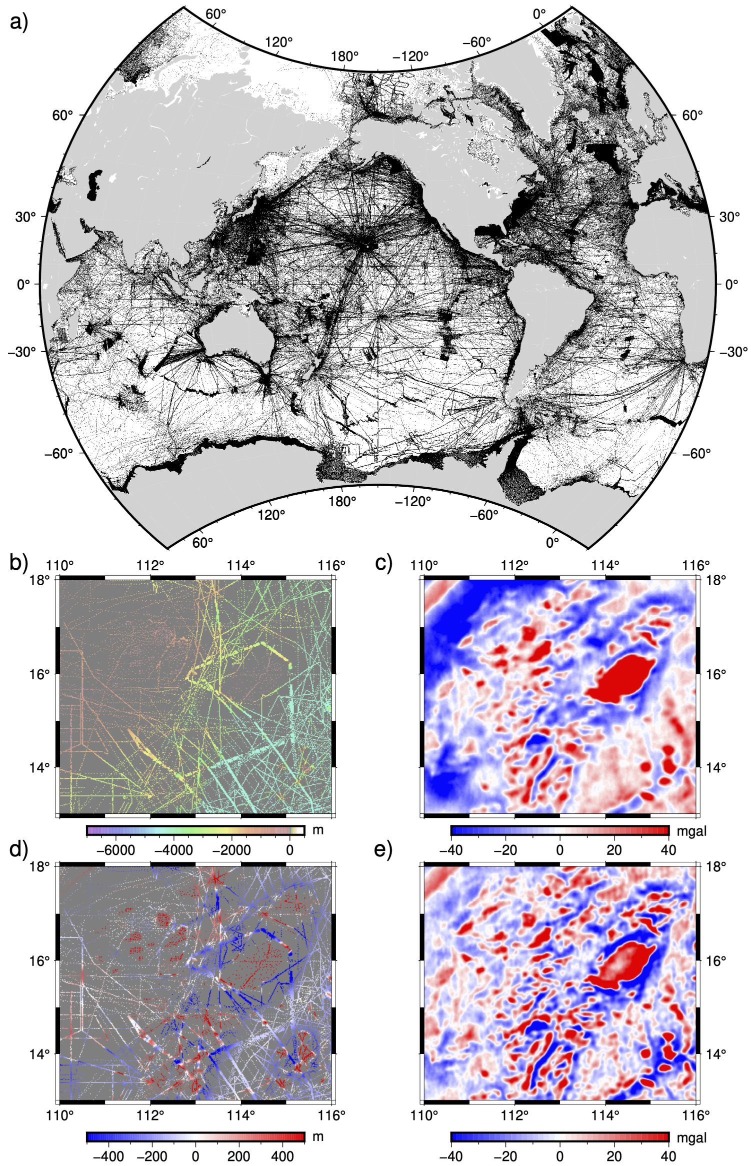


Figure 1. An overview of the datasets in map view. (a) Distribution of shipboard depth measurements in the global oceans based on publicly-available data. (b) Zoomed-in view of depth measurements in the South China Sea, colored by absolute depth. (c) Free air gravity anomaly for the same region as (b). (d) High-pass filtered depths and (e) filtered gravity anomaly–filtering described in the text.

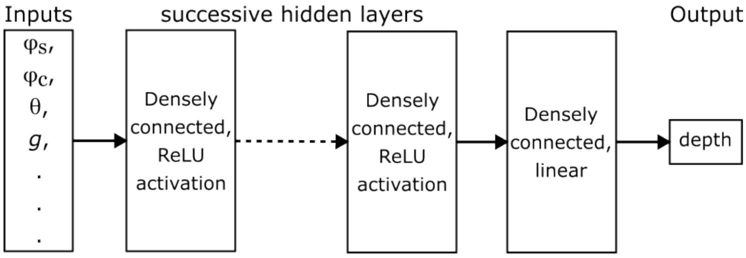


Figure 2. Schematic of the neural network architecture. A feature vector with normalized inputs is transformed by successive hidden layers to predict depth.

longitude, we avoid a discontinuity at the Greenwich Meridian in the predicted depth grid. Following this amendment, our feature vector is $[\phi_s, \phi_c, \theta, g]$.

2.1.1 Filtering depth and gravity

Since gravity and bathymetry are only correlated over certain wavelengths (Smith, 1998), it would be less-than-optimal to try and predict the absolute depth. Here we use the filtering steps established by Smith and Sandwell (1994). The depth measurements are gridded on a 1 arcminute spherical Mercator grid using continuous splines in tension (Smith & Wessel, 1990), and this grid is separated into low- and high-frequency components with a Gaussian filter with 0.5 gain at 160 km (eq. 9, (Smith & Sandwell, 1994)). The low-pass depth grid is saved. The high-pass depth is then low-pass filtered with 0.5 gain at 16 km, resulting in a 160 km - 16 km band-pass depth grid–we will call this h. The constrained points of this depth grid are extracted (Figure 1d). We are losing some high-frequency information this way in order to match the spectral content of the gravity. The gravity anomaly is high-pass filtered in the same way as the depth measurements. Finally, the high-pass filtered gravity anomaly is downward-continued to the low-pass filtered depth using a depth-dependent Wiener filter (eq. 11, (Smith & Sandwell, 1994))– we will call this g^* (Figure 1e). We will examine the effects of omitting this pre-processing step in the discussion.

2.2 Data splitting

We must split our dataset into training, validation, and testing sets. If we are to simply select 20% for testing and validation at random, then we will find that almost any given record in the testing or validation set is within 1 arc minute of a record in the training set. Marks et al. (2010), analyzing predicted depths generated by the Nettleton procedure, showed the prediction error increases with distance from constrained nodes. How strong this effect is depends on the roughness of the seafloor, but it appears the errors become decorrelated beyond a certain distance (15-30 km). In other words, records that are sufficiently close in position are not independent (Figure 3b). In practice, the training loss and validation (and testing) loss will be nearly identical, and we will not have a good idea of when the model is overfitting during training or how the model generalizes to the unmapped gaps.

By splitting the data into longitude, latitude bins and randomly selecting bins for testing and validation, we can reduce the dependence of the datasets (Figure 3c). We use a bin size of 30 arc minutes ($\sim$ 50 km at the equator) to group records, and then randomly select those bins for training, validation, and testing. This bin size could be made larger or smaller, or it could be made to vary based on prior knowledge of seafloor roughness, but it's important not to tune the bin size by model loss performance. The

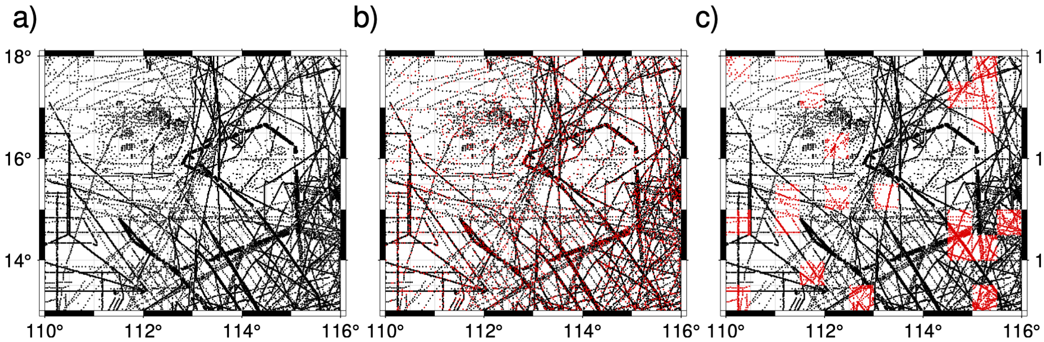


Figure 3. An example of partitioning the data. Same map area shown in Figure 1b-e. (a) The collection of depth measurements. The data must be partitioned into a training, testing, and validation set. (b) Randomly withholding 20% of the data, sampled uniformly. Withheld data are shown in red. Using this partition scheme, almost any withheld point has a nearly identical point in the training set, so the sets are not independent. (c) Sampling the data after binning into groups of 30 arc minutes. Withheld data are shown in red.

training, validation, and testing datasets comprise 31,320,896, 10,460,197, and 10,472,576 records respectively.

2.3 Model architecture and training

We use the TensorFlow software library to design and train the neural network (Abadi et al., 2015). The neural network comprises only successive densely-connected layers using a ReLU activation function (Figure 2). Input features are normalized by the mean and variance of their distribution in the training dataset. We use eight successive dense layers with 256 neurons per layer, and a final linear output layer. Model architecture can be tweaked ad nauseum, so we cannot claim this is a strictly optimal configuration, but this particular arrangement was selected because we found it performed as well as a wider but shallower model (e.g., 4 layers of 1024 neurons each) while using far fewer parameters (4e5 vs. 3e6).

We choose mean squared error (MSE) as the loss function. Each dense layer is regularized with L2 regularization ($\lambda = 0.01$). We use the Adam optimizer (Kingma & Ba, 2017) with a learning rate of 0.001. Model training proceeds until the validation loss is no longer decreasing.

2.4 Inference: generating the global predicted grid

The end goal of this model is a global grid of predicted depths. After the model is trained, predictions of h are generated on a 1 arcminute spherical mercator grid (onshore values are masked). The long wavelength depth, saved from the filtering step, are added to the predicted depth, h, to give absolute depth, d. Finally, for distribution, the predicted depth grid is "polished" with the constrained depth measurements, but this step is omitted in the following discussion and analysis.

3 Results

3.1 Base model

Using the feature vector $[\phi_s, \phi_c, \theta, g^*]$ to predict h, we achieve a training RMSE of 85 m, validation RMSE of 108 m, and testing RMSE of 109 m. Loss values are useful for comparing one trained model against another, but they are imperfect when comparing to the Nettleton method. Since no "testing" data is withheld from the Nettleton method (i.e., the prediction is tuned on all available data), we must caution the comparison of RMSE between the two methods. With that in mind, the RMSE of the Nettleton prediction and h is about 143 m. We will look more carefully at model misfit in the discussion.

3.1.1 Modeling without filtering depths and gravity

For comparison, a model trained without filtering depths and gravity anomaly performs much worse and offers no improvement over the Nettleton method. For this model, we achieve a training RMSE of 140 m, validation RMSE of 173 m, and testing RMSE of 175 m. This poor performance may be because the distribution of depth is more variable with location. For example, where the regional depth is 6000 m, the mean depth will be near 6000 m, and similar for a regional depth of 1000 m. By high-pass filtering the depth, the overall variance of the data are reduced. Omitting the low-pass filter at 16 km also contributes to the worse performance. The Nettleton method RMSE above is for the band-pass filtered data, h. If the short wavelengths are included, the Nettleton RMSE is about 165 m.

Alternatively, we can high-pass filter depth and gravity but omit the low-pass filter at 16 km–in fact, we may desire to do this so we don't lose short-wavelength details. For this model, we achieve a training RMSE of 124 m, validation RMSE of 141 m, and testing RMSE of 142 m. We can't compare these directly to the base model since the target quantities are different. Instead we can evaluate the RMSE of the base model prediction and the high-pass depth. In this case, testing RMSE values are nearly identical for the two models, suggesting the trained models are similar and the greater loss reflects the greater variance of the target data.

3.2 Added features

It would seem that adding features from other global grids would be an easy way to decrease model loss and improve performance. For example, the spreading rate at the time crust is created is known to affect the roughness of bathymetry (Small & Sandwell, 1994). Crustal age and sediment cover will also influence the correlation of gravity and bathymetry (Smith & Sandwell, 1994). We use ϕ, θ to sample grids of crustal age, paleo-spreading rate (Seton et al., 2020), and sediment thickness (Whittaker et al., 2013), add these to the feature vector, and train a new model. In practice, there are problems with using these features.

Firstly, these grids have many regions of missing data, and the missing values must be handled somehow. Since a key purpose of this model is prediction on a global grid, grid cells with missing values can't simply be thrown out. We tested different schemes to replacing missing values: replacement with the mean feature value; replacement with mean feature value plus an additional boolean feature indicating replacement; and filling missing values with nearest neighbor interpolation of the feature grid. In our attempts to use crustal age, paleo-spreading rate, and sediment thickness as features, validation and testing loss are not improved (nor are they improved by any one such feature), and in fact model loss is worse with the additional features. In addition, at inference time, sharp discontinuities in the feature grids get mapped to the prediction grid creating unwanted artifacts.

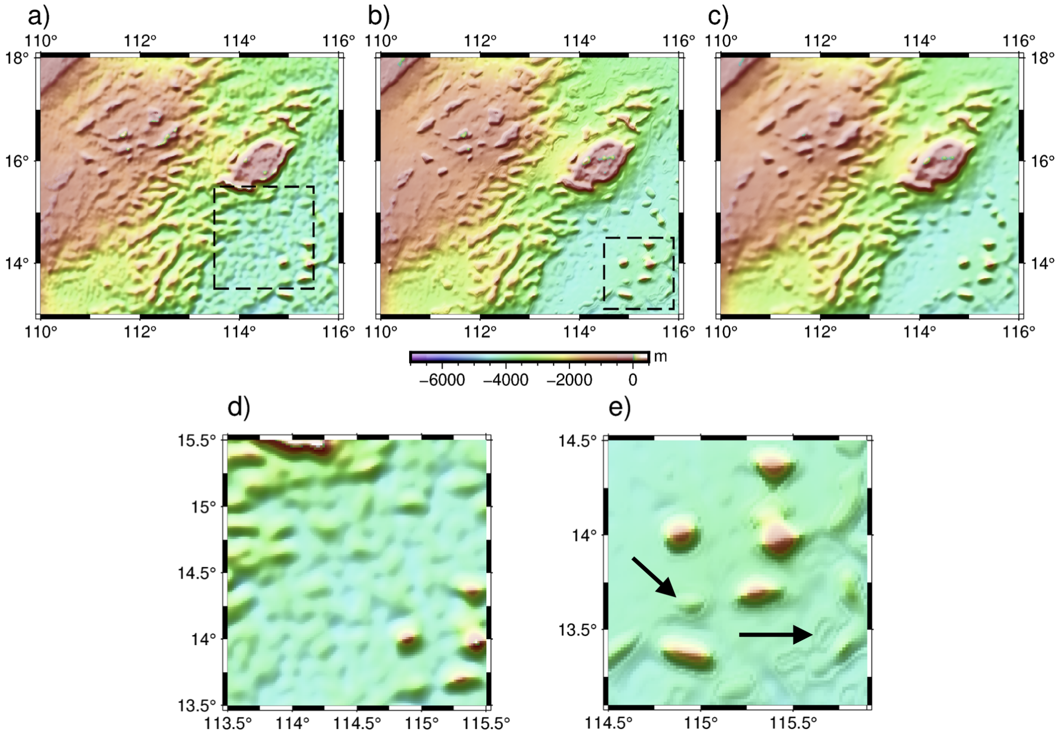


Figure 4. Model-predicted depths with long-wavelength depth added. The same map area shown in Figure 1b-e. (a) Nettleton method; (b) Raw DNN-predicted grid; (c) DNN-predicted grid, low-pass filtered at 12 km. (d) Example of the "orange peel" texture in the Nettleton prediction, boxed region in (a). (e) Example of "hallucinations" in DNN prediction, boxed region in (b).

Other gravity-related quantities such as deflection of the vertical and vertical gravity gradient (VGG) may be good features to use, but they may only be redundant. We found that adding VGG as a feature (Sandwell et al., 2021) improves model training RMSE only slightly, and it slightly worsens validation and testing RMSE. This model has a training RMSE of 85 m, a validation RMSE of 109 m, and a testing RMSE of 110 m. Since there is no improvement on the base model, we will consider that the preferred model and refer to it as simply the "DNN" model in the following discussion.

4 Discussion

4.1 Generating a predicted depth grid

After training, we generate model predicted depths on a global 1 min mercator grid. One problem that results is short wavelength artifacts or "hallucinations" (Figure 4e). These hallucinations typically occur with wavelengths shorter than the 16 km wavelength filter that was applied to the original gravity and bathymetry, so they must be a product of the DNN training. We can reduce these with regularization during training, but not completely nor in a deterministic way. For the distributed predicted depth grid, we apply a low pass filter with 0.5 gain at 16 km to remove these hallucinations. This post-inference filtering method does not weaken model results. In fact, error metrics are very slightly improved, and the prediction RMSE of h after filtering is 107 m for the testing dataset. We use the predictions on the filtered grid in the following discussion.

4.2 Comparison to Nettleton model

While not quantitative, it's important to visually inspect the DNN predicted depth grid and make comparisons to the Nettleton grid (Figure 4). The most obvious qualitative difference is in continental margin areas or areas of relatively smooth seafloor. In these areas, the Nettleton predicted bathymetry has a rough "orange peel" texture (Figure 4d), an artifact of downward continuation of noisy gravity data. This type of seafloor is smoother in the DNN prediction.

Using all available depth measurements (not partitioned for training/testing), we compare the error distribution of the Nettleton method and the DNN method. These results are shown in Figure 5. For the Nettleton method, the predicted depth is within 68 m for 50% of points and within 168 m for 80% of points. For the DNN method, these percentiles are 45 m and 128 m, respectively. Additionally, the mean error has been reduced from 13 m for the Nettleton to 3 m for the DNN, indicating a less-biased estimate. Figure 5b shows the distribution of absolute model error in the southern oceans. Overall, the spatial patterns of misfit are similar for the two models. At this scale, the noticeable differences are found nearer to land–e.g., the West Antarctic Peninsula, Chile, Australia–where the DNN model shows clear improvements over the Nettleton.

4.2.1 BODC data

Because the Nettleton prediction is tuned using all available data, we don't have a concrete idea of how well it generalizes to unseen areas of seafloor. It is useful to reserve a dataset that is not used in either model's construction. To compare the performance of the Nettleton and the DNN model, we used a collection of depth measurements from 279 cruises from the British Oceanographic Data Centre (BODC) that are not yet incorporated into the prediction model. The raw data are decimated to 15 sec, and measurements that overlap with data already in the prediction dataset are removed. We did not thoroughly inspect the BODC data for erroneous measurements, so measurements that differ from either predicted grid by more than 2000 m are removed. In total, there are 6,242,414 points. The Nettleton prediction has an RMSE of 150 m for the BODC dataset. The final DNN model has an RMSE of 104 m.

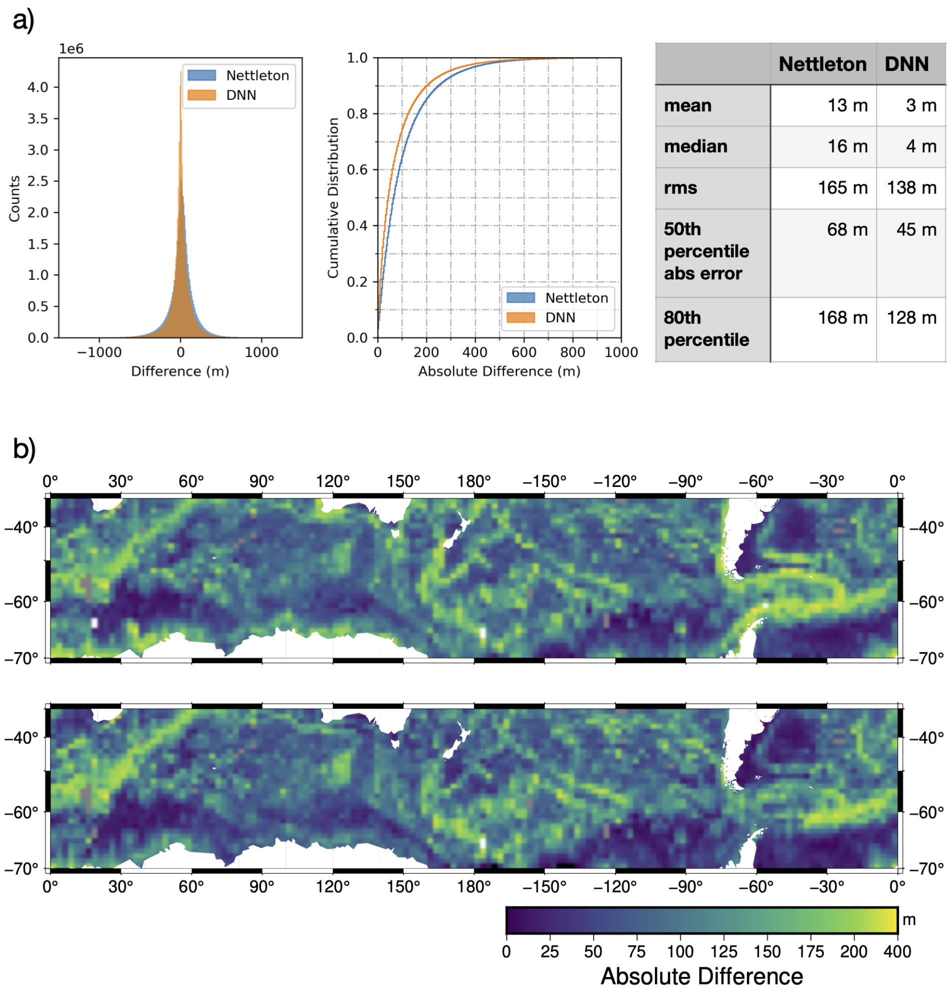
If we restrict the analysis spatially to the highest concentration of measurements (80% of the data are around the British Isles), the Nettleton RMSE is 73 m and the DNN model RMSE is 62 m–much lower than testing RMSE (Figure 6). This almost certainly reflects the proximity of these data to those in our training set. However, we see from the error distribution that the slight bias in the Nettleton prediction is not present in the DNN prediction.

4.3 Potential for improvements

Our model is a simple implementation of a neural network to predict depth globally, and we have shown its clear improvement over the Nettleton method. Yet, there are many possible directions for improvement depending on one's objectives. Expansion of the training dataset, modifications of model architecture, or a multi-regional approach to the problem all offer potential to improve on our model.

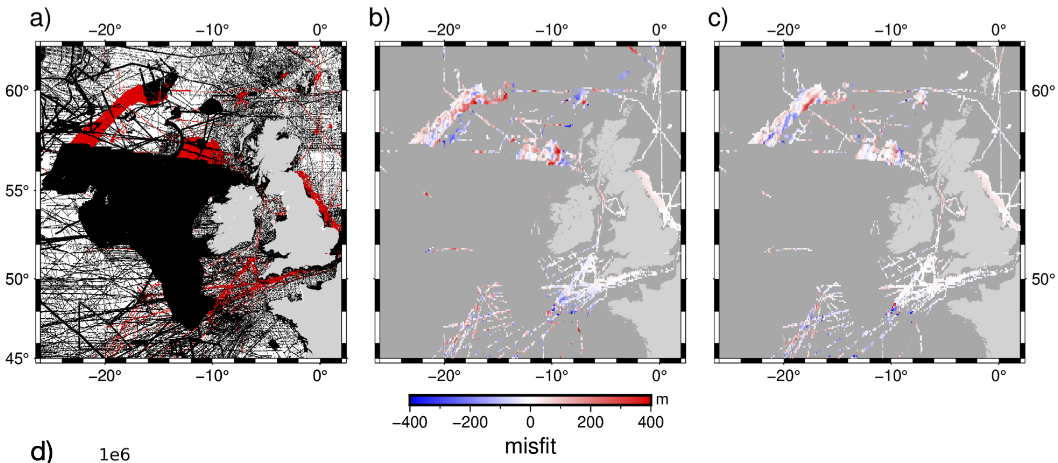
If coverage of publicly available bathymetry compilations continues to improve as it has in recent years–and it likely will (Mayer et al., 2018)–model predictions will clearly improve (this would be true of any model). Low-resolution data in remote regions, which can be collected by autonomous vehicles, will likely offer the greatest benefit in our model approach.

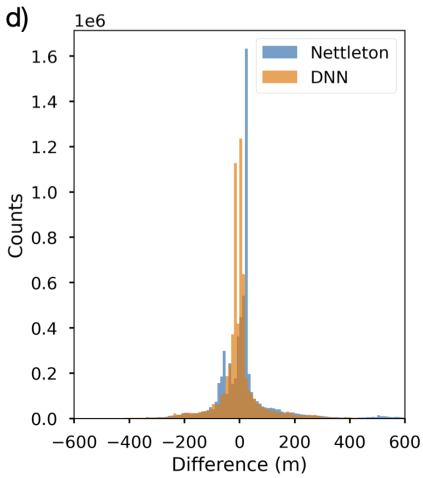
We haven't made use of high-resolution multibeam data in our model, and we do not aim to predict features at such resolution. Upward continuation of gravity anoma-



	Nettleton	DNN
mean	13 m	3 m
median	16 m	4 m
rms	165 m	138 m
50th percentile abs error	68 m	45 m
80th percentile	168 m	128 m

Figure 5. Comparison of Nettleton and DNN models by prediction error. (a) Distribution of prediction error for all 1 arc minute data (N=52,253,670). (b) Average absolute difference (prediction - measurement) for Nettleton (upper) and DNN (lower) models.





	All		**British Isles**	
	Nettleton	**DNN**	**Nettleton**	**DNN**
mean	22 m	-3 m	12 m	-1 m
median	19 m	-3 m	20 m	-1 m
rms	150 m	104 m	73 m	62 m
50th percentile abs error	28 m	16 m	25 m	14 m
80th percentile	94 m	71 m	60 m	40 m

Figure 6. Nettleton and DNN model predictions for withheld BODC data. (a) Depth measurements in the full dataset shown in black, measurements from the disjoint BODC dataset shown in red. Misfit of BODC data for (b) Nettleton predicted depth and (c) neural network predicted depth. (d) Distribution of BODC data misfit for Nettleton predicted depth and neural network predicted depth.

lies limits the resolution of gravity from satellite altimetry to a length scale of about π times the regional depth (e.g. Smith & Sandwell, 2004), so it is not possible to realistically predict depth from only gravity (and its derivatives) at such scales. An approach using convolutional neural networks, as demonstrated by Annan and Wan (2022), may successfully learn from higher resolution bathymetry in regional settings.

A prediction model trained on regional data will, everything else equal, perform better in that region than a model trained on global data. Moran et al. (2022) identified regions where various learning algorithms might preferentially excel. This suggests a global model should alternatively be constructed from a suite of regional models. A particular case where such a multi-regional model would excel is in predicting higher resolution depth in areas where that is realistic. Susa (2022) showed such an approach to predicting depth in near-coastal regions. In this setting, altimetric ranging and gravity accuracy suffer from land contamination (Raney & Phalippou, 2011), and visible spectra may be correlated with bathymetry, making this an ideal case for alternative depth prediction.

5 Conclusions

1. Using a large collection of depth measurements and satellite-derived gravity anomalies, we trained a deep neural network to predict seafloor depth.
2. We find that applying filters (described by Smith and Sandwell (1994)) to bathymetry and gravity before training is necessary for a good result, and conforms the data more closely with the assumption of identical distributions.
3. When dealing with sparse heterogeneous sampling, the training-testing split must be treated carefully. If the training and testing split is a random partition at the same resolution as the data, the training and testing sets are not independent, and model misfit results will be too optimistic.
4. Adding features sampled from geologic grids–crustal age, paleo-spreading rate, sediment thickness–do not improve model results.
5. Our preferred DNN-predicted model improves on the results of the Nettleton procedure, lowering the prediction RMSE from 165 m to 138 m.
6. While improvements will be made with additional depth measurement data, higher resolution predictions are limited by the upward continuation of gravity, so likely shouldn't be attempted with this method.

Open Research Section

Jupyter notebooks and data files to reproduce the predicted depth model, as well as the final predicted depth model used in the analysis, are available at `https://doi.org/10.5281/zenodo.8029925` (Harper & Sandwell, 2023).

Acknowledgments

We thank Ross Parnell-Turner, Dave May, Jose Restrepo, and Jeff Gee for their suggestions and comments on a preliminary version of the manuscript. This work was supported by the NASA SWOT program (80NSSC20K1138), a NASA FINNEST fellowship (80NSSC20K1616), and the Office of Naval Research (N00014-17-1-2866).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., . . . Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Retrieved from `https://www.tensorflow.org/` (Software available from tensorflow.org)

Annan, R. F., & Wan, X. (2022, 10). Recovering bathymetry of the gulf of guinea using altimetry-derived gravity field products combined via convolutional neural network. *Surveys in Geophysics*, *43*, 1541-1561. doi: 10.1007/s10712-022-09720-5

Becker, J. J., Sandwell, D. T., Smith, W. H., Braud, J., Binder, B., Depner, J., ... Weatherall, P. (2009, 10). Global bathymetry and elevation data at 30 arc seconds resolution: Srtm30_plus. *Marine Geodesy*, *32*, 355-371. doi: 10.1080/01490410903297766

GEBCO compilation group. (2023). *Gebco 2023 grid.* doi: 10.5285/f98b053b-0cbc-6c23-e053-6c86abc0af7b

Harper, H., & Sandwell, D. (2023, June). *Global predicted bathymetry using neural networks.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.8029925` doi: 10.5281/zenodo.8029925

Kingma, D. P., & Ba, J. (2017). *Adam: A method for stochastic optimization.*

Marks, K. M., Smith, W. H. F., & Sandwell, D. T. (2010). Evolution of errors in the altimetric bathymetry model used by google earth and gebco. *Marine Geophysical Research*, *31*, 223-238. doi: 10.1007/s11001-010-9102-0

Mayer, L., Jakobsson, M., Allen, G., Dorschel, B., Falconer, R., Ferrini, V., ... Weatherall, P. (2018). The nippon foundation-gebco seabed 2030 project: The quest to see the world's oceans completely mapped by 2030. *Geosciences (Switzerland)*, *8*. doi: 10.3390/geosciences8020063

Moran, N., Stringer, B., Lin, B., & Hoque, M. T. (2022). Machine learning model selection for predicting bathymetry. *Deep-Sea Research Part I: Oceanographic Research Papers*, *185*, 103788. doi: 10.1016/j.dsr.2022.103788

Nettleton, L. L. (1939). Determination of density for reduction of gravimeter observations. *Geophysics*, *4*.

Parker, R. (1972). The rapid calculation of potential anomalies. *Geophysical Journal of the Royal Astronomical Society*, *31*, 447-455. doi: 10.1111/j.1365-246X.1973.tb06513.x

Raney, R. K., & Phalippou, L. (2011). The future of coastal altimetry. In S. Vignudelli, A. G. Kostianoy, P. Cipollini, & J. Benveniste (Eds.), *Coastal altimetry* (pp. 535–560). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-12796-0_20

Sandwell, D. T., Harper, H., Tozer, B., & Smith, W. H. F. (2021). Gravity field recovery from geodetic altimeter missions. *Advances in Space Research*, *68*(2), 1059–1072.

Seton, M., Müller, R. D., Zahirovic, S., Williams, S., Wright, N. M., Cannon, J., ... McGirr, R. (2020). A global data set of present-day oceanic crustal age and seafloor spreading parameters. *Geochemistry, Geophysics, Geosystems*, *21*, 1-15. doi: 10.1029/2020GC009214

Small, C., & Sandwell, D. T. (1994). Imaging mid-ocean ridge transitions with satellite gravity. *Geology*, *22*, 123-126.

Smith, W. H. F. (1998). Seafloor tectonic fabric from satellite altimetry. *Annual Review of Earth and Planetary Sciences*, *26*, 697-738.

Smith, W. H. F., & Sandwell, D. T. (1994). Bathymetric prediction from dense satellite altimetry and sparse shipboard bathymetry. *Journal of Geophysical Research*, *99*. doi: 10.1029/94jb00988

Smith, W. H. F., & Sandwell, D. T. (1997). Global sea floor topography from satellite altimetry and ship depth soundings. *Science*, *277*, 1956-1962. doi: 10.1126/science.277.5334.1956

Smith, W. H. F., & Sandwell, D. T. (2004). Conventional bathymetry, bathymetry from space, and geodetic altimetry. *Oceanography*, *17*, 8-23. doi: 10.5670/oceanog.2004.63

Smith, W. H. F., & Wessel, P. (1990). Gridding with continuous curvature splines in tension. *Geophysics*, *55*(3), 293–305.

Susa, T. (2022). Satellite derived bathymetry with sentinel-2 imagery: Comparing traditional techniques with advanced methods and machine learning ensemble models. *Marine Geodesy*, *0*, 1-20. doi: 10.1080/01490419.2022.2064572

Tozer, B., Sandwell, D. T., Smith, W. H. F., Olson, C., Beale, J. R., & Wessel, P. (2019). Global bathymetry and topography at 15 arc sec: Srtm15+. *Earth and Space Science*, *6*, 1847-1864. doi: 10.1029/2019EA000658

Wan, X., Annan, R. F., & Ziggah, Y. Y. (2023, 2). Altimetry-derived gravity gradients using spectral method and their performance in bathymetry inversion using back-propagation neural network. *Journal of Geophysical Research: Solid Earth*, *128*. doi: 10.1029/2022JB025785

Watts, A. B. (2001). *Isostasy and flexure of the lithosphere.* Cambridge University Press.

Whittaker, J. M., Goncharov, A., Williams, S. E., Müller, R. D., & Leitchenkov, G. (2013). Global sediment thickness data set updated for the australian-antarctic southern ocean. *Geochemistry, Geophysics, Geosystems*, *14*(8), 3297-3305. doi: https://doi.org/10.1002/ggge.20181