

Signal Separation in Global, Temporal Gravity Data

Betty Heller-Kaikov
Technical University of Munich
Germany
betty.heller@tum.de

Roland Pail
Technical University of Munich
Germany
roland.pail@tum.de

Martin Werner
Technical University of Munich
Germany
martin.werner@tum.de

ABSTRACT

Satellite gravity data such as provided by the Gravity Recovery and Climate Experiment (GRACE) and its follow-up mission contain valuable information on all geophysical processes that involve a mass redistribution in the Earth system. However, as gravity is an integral quantity, only the sum of all signal components can be measured. In order to exploit gravity data for improving the geophysical understanding of the underlying mass redistribution processes, an algorithm to separate the superimposed signals is needed. We present two methods that solve the signal separation problem in global, spatial-temporal gravity data based on prior knowledge on the characteristic behavior of the individual signal components: A principal component analysis-based method is compared to a conditional generative adversarial network that has been originally developed for image-to-image translation tasks. Both methods are tested on synthetic model data in a closed-loop setup, and are shown to successfully accomplish the task of separating gravity signals related to atmospheric and oceanic processes from signals caused by processes in the continental hydrosphere. The two methods show a comparable level of prediction errors. For both methods, the preprocessing of the gravity signals before the signal separation step, more precisely, the choice of the gravity field functional, has a considerable impact on the predictive performance. All in all, we emphasize the potential of neural network algorithms for signal separation in spatial-temporal geodetic data, and suggest a more specific tuning of the algorithm to the task of interest.

CCS CONCEPTS

• Applied computing → Environmental sciences; Earth and atmospheric sciences.

KEYWORDS

satellite gravimetry, signal separation, spatial-temporal data, neural networks

ACM Reference Format:

Betty Heller-Kaikov, Roland Pail, and Martin Werner. 2023. Signal Separation in Global, Temporal Gravity Data. In *6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GeoAI '23)*, November 13, 2023, Hamburg, Germany. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3615886.3627743>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

GeoAI '23, November 13, 2023, Hamburg, Germany

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 979-8-4007-0348-5/23/11...\$15.00
<https://doi.org/10.1145/3615886.3627743>

1 INTRODUCTION

The Earth's gravity field changes in time due to mass redistribution processes in the hydrosphere, the atmosphere, the oceans, the cryosphere and the solid Earth. Observing these temporal gravity variations, e.g. by dedicated satellite gravity missions, provides valuable insight into the underlying geophysical processes causing the mass changes. The Gravity Recovery and Climate Experiment (GRACE, mission period 2002-2017, [18]) and its follow-up mission GRACE-Follow on (GRACE-FO, mission in orbit since 2018, [4], [10]) provide global datasets of monthly to sub-monthly variations of the Earth's gravity field at a spatial resolution of a few 100 km. In order to extend the gravity data time series beyond the GRACE-FO mission lifetime, as well as to increase the temporal and spatial resolution of the dataset, as future gravity mission concept the Mass Change And Geoscience International Constellation (MAGIC) is currently developed and designed ([11], [7]).

While the data of the GRACE(-FO) mission are used for a multitude of scientific applications (see e.g. [2] for a recent review), one major challenge in the processing of gravity data that is still not finally solved is the problem of signal separation: Since gravity is an integral quantity, only the sum of all mass variations can be measured. A separation of the total signal into the individual sub-signals and thus the assignment to their original sources needs to be performed as a part of the data processing.

In general, there exist several approaches for signal separation of spatial-temporal geodetic data. The most commonly used method is to subtract all signal components which are not of interest from the data using external geophysical models for these components. This approach has the disadvantage that model errors are directly

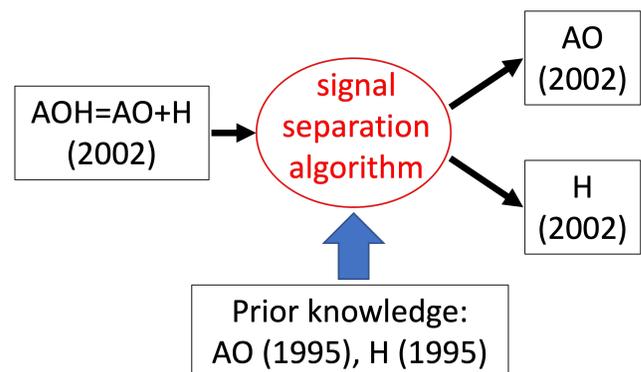


Figure 1: The algorithm to be developed is able to separate spatial-temporal gravity signals related to atmosphere and ocean (AO) and hydrological (H) processes from their sum AOH.

propagated to the target signal, which can lead to significant error contributions in the estimated signal of interest (e.g. [14], [19]).

A further approach is the combination and joint inversion of the data with other datasets that have a different functional relationship to the mass change phenomenon of interest and are therefore complementing the information of GRACE(-FO) (e.g. [20], [24]). However, this approach is not applicable if complementary datasets do not exist for the considered process, which is often the case.

Another group of signal separation methods that avoids the introduction of model errors are algorithms exploiting the statistically differing behavior of the signal components to be separated. Methods based on the principal component analysis (PCA; [13], [15], [6]) derive typical spatial patterns for the individual signal components from auxiliary (model or observational) data and use them as prior knowledge to extract the signal components from their measured sum. Independent component analysis (ICA; [5], [1], [16]) methods, in contrast, perform a form of blind source separation, assuming statistical independence of signals coming from independent physical processes.

While the PCA methods recognize the individual signal components based on their main patterns in space domain, in gravity data, typical patterns also exist in the temporal domain. Therefore, we attempt to exploit as much prior knowledge on the individual signal components as possible, by performing a supervised training of a neural network architecture using model data for the individual signal components as training data.

In the area of training neural networks on signal separation tasks, we have found previous work on time series (audio) data (e.g. [22], [12]) or 2d-image data (e.g. [17]). Selecting a suitable architecture for signal separation of spatial-temporal data, however, does not seem straightforward. In our paper, we investigate the applicability of a typical image processing algorithm for our purpose. As it is an application-independent algorithm that has already been successfully applied to a large variety of tasks, we select the conditional generative adversarial network pix2pix [9] as representative of such an algorithm. As a reference, we compare the obtained results to corresponding results using the PCA-based method from [15].

In the present paper, we limit ourselves to these two statistical methods. The other above-described signal separation strategies (subtraction of non-target signals using external geophysical models; joint inversion of multiple complementary datasets) are employed in real data processing, but cannot be transferred directly to the here-considered simulation test case.

As shown by Fig. 1, we set up a synthetic closed-loop test environment, where we investigate as a representative case study the task of separating spatial-temporal gravity signals caused by processes in the atmosphere and oceans (AO) and hydrospheric signals (H) from their sum AOH. The AO and H signals are chosen as sub-signals as they represent the largest-amplitude signal parts in temporal gravity data (except of glaciated areas). The A and O components are not separated further, as due to the inverted barometer effect [23], gravity signals due to mass variations in the atmosphere and oceans are not perfectly separable in oceanic regions.

Prior knowledge on the characteristic signal behavior is introduced in the form of model data of the year 1995, while the corresponding model data of the year 2002 are used for testing. While this is a highly simplified setup which is not fully representative

of the signal separation task in real satellite gravity data, which include a much higher number of signal components as well as noise, we consider this setup as a first step in the algorithm development and performance assessment, which in future work will need to become refined towards a higher complexity.

Our paper is structured as follows: Section 2 introduces the ESA Earth System Model, which serves as dataset in this study. In section 3, both the PCA-based as well as the pix2pix neural network algorithms are described. Section 4 contains selected experiments to evaluate the performance of the two investigated families of algorithms. Finally, we conclude our study in section 5 and give a brief outlook.

2 THE ESA EARTH SYSTEM MODEL DATASET

For the development and testing of signal separation methods, we work with closed-loop simulations, as they allow to assess the performance of the investigated methods by computing their prediction error, which is the difference between the predicted and the respective true signal components. As model for the individual gravity signal components, we are using the updated Earth System Model of ESA (ESM, [3]). The latter is a synthetic model for variations in the Earth's gravity field induced by realistic mass variations caused by geophysical processes in the atmosphere (A), the oceans (O), the continental hydrosphere (H), the cryosphere (I, for ice), and the solid Earth (S). The model spans the period 1995 to 2006 and gives the respective gravity variations parameterized in terms of spherical harmonic (SH) coefficients for the temporal deviations of the gravity potential from its 11-year mean. The maximum SH degree and order is 180, representing a spatial resolution of about 110 km. The temporal resolution is 6 h, such that for each model year, 1460 (depending on the number of days) sets of SH coefficients are provided. As described in section 1, for our study we use the model data of the years 1995 and 2002, and consider the sum AO of the A and O model components as well as the H model component.

The full preprocessing flow starting from the SH coefficient sets is visualized in Fig. 2. Each set of SH coefficients is expanded to a spatial (363-by-725 latitude-longitude) grid of equivalent water heights (EWHs) according to [21], or as an alternative, of geoid heights according to [8]. Both quantities are scalar-valued and represent functionals describing the time- and position-dependent strength of the Earth's gravity field. The main mathematical difference of these two functionals (disregarding their physical meaning) is that for the computation of EWHs, the SH coefficients are scaled according to their SH degree, leading to an amplification of large-SH degree (i.e. small-scale) signal components in the resulting data grids.

In the next step, the spatial grids are concatenated to form a 363-by-725-by-1460 latitude-longitude-time data matrix containing the data of one model year, and interpolated along the longitude and time axes to 256 samples each. The latter step is performed in order to match the data format expected by the pix2pix cGAN algorithm [9], and represents a down-sampling from 55 km to about 156 km in the longitude direction, and from 6 h to about 34 h in the time direction. Also, the data is normalized to values between -1 and 1 using a fixed normalization factor.

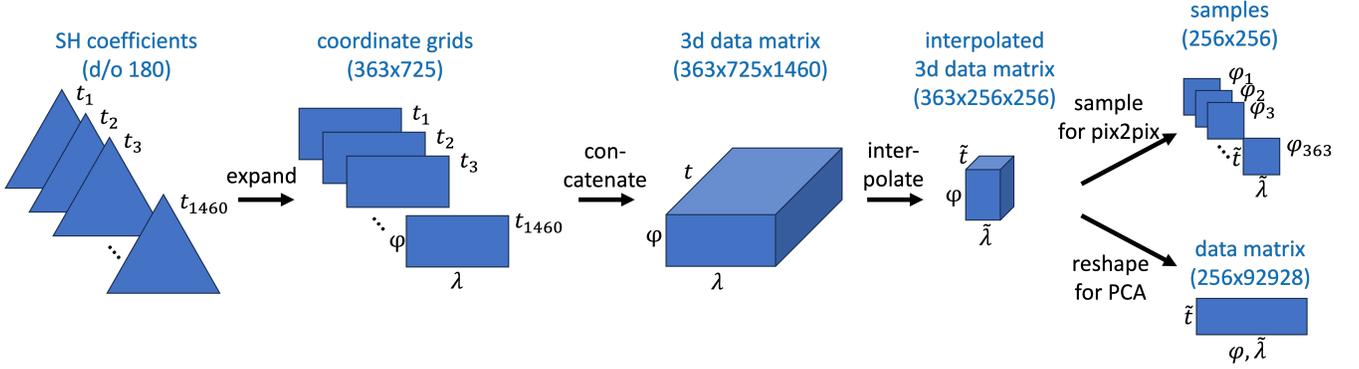


Figure 2: Preprocessing flow applied to each data year of the considered ESM components (A, O, H), starting from the d/o 180 SH coefficient sets each describing a 6-hour mean anomalous A, O or H gravity potential. The data preparation only splits up to the data formats required by the two signal separation methods (cf. section 3) at the very end.

We note that the described down-sampling is a limitation that needs to be improved in a future neural network architecture for signal separation. However, our goal is to test the original pix2pix code for our question of interest. Therefore, we present respective results using the interpolated data in this paper.

Only in the very last step visualized by Fig. 2, the data is brought to the formats expected by the two methods: As the pix2pix method (cf. section 3.2) operates on 2-dimensional square images as training or testing data samples, 256-by-256 time-longitude slices are extracted from the interpolated 363-by-256-by-256 data matrix. This results in 363 2-dimensional data samples, each containing the data values of one fixed latitude value, i.e. representing the temporal evolution of one pixel row in the original latitude-longitude coordinate grid. For the PCA method (cf. section 3.1), the interpolated 363-by-256-by-256 data matrix is reshaped to one large 2-dimensional matrix containing the complete global, 1-year dataset. In this 256-by-92928 data matrix, each row contains the data values of one point in time, with the columns corresponding to all latitude-longitude pairs of the complete coordinate grid.

3 METHODS

3.1 Principal Component Analysis (PCA)

The idea of principal component analysis (PCA, [13]) is to extract typical spatial patterns for each of the two considered signal components (here: AO and H) and to use these patterns, also called empirical orthogonal functions (EOFs), as basis vectors spanning two signal component subspaces. In the following, we briefly outline the method as used in our application.

As a first step, each component is considered separately to find the EOFs associated with its spatial-temporal signal via PCA:

$$\begin{aligned} AO_{1995}(t, p) &= \sum_i c_{AO,i}(t) EOF_{AO,1995,i}(p) \\ H_{1995}(t, p) &= \sum_i c_{H,i}(t) EOF_{H,1995,i}(p), \end{aligned} \quad (1)$$

where AO_{1995} and H_{1995} are the 256-by-92928 time-position gridded datasets of the AO and H signals of the year 1995 (see

Fig. 2, bottom right). t denotes time, p the latitude- and longitude-dependent position on the global coordinate grid, $c_{\{AO,H\},i}$ the principal components representing the temporal evolution of the individual modes, and $EOF_{\{AO,H\},i}$ the above-mentioned spatial patterns of the AO and H signals, respectively. As the used dataset comprises 256 time steps, a total of 256 linearly independent EOFs is computed. In Eq. (1), the EOFs are sorted by decreasing amount of input signal energy covered.

This first step of computing the EOFs corresponds to building the signal separation model based on prior information of the year 1995. In the following second and third step, this PCA separation model is applied to the ESM data of the year 2002. The second step consists of expressing the full signal AOH as linear combination of the (before-computed) EOFs of the individual signals:

$$\begin{aligned} AOH_{2002}(t, p) &\approx \sum_{i=1}^{N_{AO}} \hat{c}_{AO,i}(t) EOF_{AO,1995,i}(p) \\ &+ \sum_{i=1}^{N_H} \hat{c}_{H,i}(t) EOF_{H,1995,i}(p) \end{aligned} \quad (2)$$

and subsequently computing the time-variable coefficients $\hat{c}_{AO,i}(t)$ and $\hat{c}_{H,i}(t)$ via Least-Squares-Adjustment (LSA).

In the third step, the individual components \widehat{AO}_{2002} and \widehat{H}_{2002} are retrieved using the coefficients \hat{c} computed in the second step:

$$\begin{aligned} \widehat{AO}_{2002}(t, p) &= \sum_{i=1}^{N_{AO}} \hat{c}_{AO,i}(t) EOF_{AO,1995,i}(p) \\ \widehat{H}_{2002}(t, p) &= \sum_{i=1}^{N_H} \hat{c}_{H,i}(t) EOF_{H,1995,i}(p) \end{aligned} \quad (3)$$

The matrices $\widehat{AO}_{2002}(t, p)$ and $\widehat{H}_{2002}(t, p)$ are subsequently reshaped to 363-by-256-by-256 latitude-longitude-time matrices for the predicted AO and H signals, respectively.

The number of EOFs, N_{AO} and N_H , used in Eqs. (2) and (3) could e.g. be tuned by iterating over the second and third of the above-described steps and choosing the values resulting in the smallest

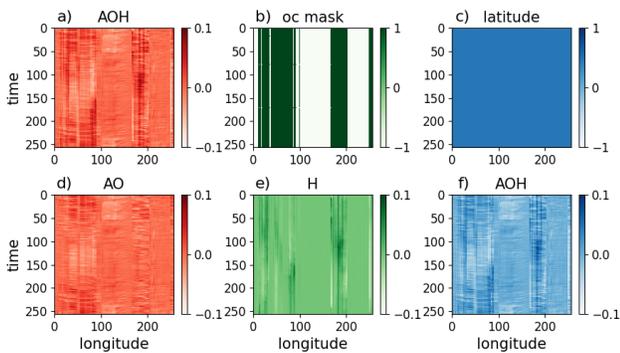


Figure 3: One training sample used for the pix2pix method. Panels a-c (d-f) show the 3 color channels of the input (target) image.

prediction errors $\widehat{AO}_{2002} - AO_{2002}$ and $\widehat{H}_{2002} - H_{2002}$. In our study, we use all modes, i.e. fix $N_{AO} = N_H = 256$, which minimizes the prediction errors on the test dataset in the case of EWH data. In general, reducing the number of modes in Eqs. (2) and (3) has the benefit of reducing the overlap between the subspaces spanned by the AO and H EOFs, but the drawback that the signal variability covered by the two subspaces is reduced.

3.2 Conditional Generative Adversarial Network

As neural network method, we use the conditional generative adversarial network (cGAN) pix2pix [9]. It consists of two parts: The generator has the task to translate a 256-by-256 3-channel input image to an image of similar dimensions, while the discriminator is trained to distinguish the images produced by the generator from true images, when given an input-output image pair. Both network parts are simultaneously trained, in order to achieve a balance between the generator’s ability to fool the discriminator and the discriminator performance.

To transform our data to the format expected by the pix2pix algorithm, we use the 256-by-256 time-longitude data samples (cf. Fig. 2 top right), and fill the sum of the considered signals, $AO+H=AOH$, in the first color channel of the input image, and the individual signal components AO and H in the first and second color channel of the output image, respectively. Figure 3 visualizes the color channels of one input-output image pair used for the pix2pix training.

The two remaining color channels of the input image are filled with supplementary information that might be useful for the training: In the second channel (Fig. 3 b), values of -1 and 1 are given for oceanic and continental pixel points, respectively, and the pixels of the third channel are filled with the (normalized, constant) latitude value of the sample. The remaining color channel of the output image is filled with the same values as the first input channel. We note that by the latter, we introduce an additional task to be learned by the net, which is to remember the AOH signal. In tests not shown in this paper, this additional task seemed to be beneficial for the main task (being the correct prediction of the AO and H signal components). However, we emphasize that the performance of the

trained nets is evaluated on the predicted first and second output channels, while the third output channel is completely disregarded.

The architecture as well as the loss functions governing the training process are not changed compared to the original pix2pix algorithm. Additional constraints concerning the signal separation task (such as e.g. a loss term forcing the sum of the predicted components AO and H being equal to the AOH input signal) are not introduced, as our objective is to evaluate the performance of the un-modified pix2pix algorithm for a signal separation task. For the batch size, we test values of 1 and 64, for the learning rate, we use $1e-4$, $5e-5$ or $1e-5$, and a fixed number of 1000 training epochs is used for all separation models.

As described in section 2, the ESM data of the year 1995 (2002) are used to train (test) the pix2pix separation models. The testing consists of successively running all samples of the test dataset through the trained separation model, and subsequently re-sorting the output data to build a 363-by-256-by-256 latitude-longitude-time matrix for the predicted AO and H signals each, covering the complete (global, 1-year) dataset.

4 EXPERIMENTS

4.1 Experimental Setup

The signal separation task visualized by Fig. 1 is solved based on the same data derived from the ESA ESM (cf. section 2) using the two methods introduced in section 3. Concretely, we consider the following signal separation models, which differ by the used gravity field functional as well as, in the case of the pix2pix models, by the used batch size (b.s.) and learning rate (l.r.):

- *pix2pix* ($E, 1, 5e-5$): EWH model data; b.s.=1, l.r.= $5e-5$
- *pix2pix* ($E, 64, 5e-5$): EWH model data; b.s.=64, l.r.= $5e-5$
- *pix2pix* ($E, 1, 1e-4$): EWH model data; b.s.=1, l.r.= $1e-4$
- *pix2pix* ($E, 1, 1e-5$): EWH model data; b.s.=1, l.r.= $1e-5$
- *pix2pix* ($G, 1, 5e-5$): geoid height model data; b.s.=1, l.r.= $5e-5$
- *PCA* (E): EWH model data
- *PCA* (G): geoid height model data

These signal separation models were selected in order to perform the following three experiments, the results of which are presented and discussed in the following sections 4.2, 4.3 and 4.4:

- **Experiment (1):** Compare *pix2pix* ($E, 1, 5e-5$), ($E, 64, 5e-5$), ($E, 1, 1e-4$) and ($E, 1, 1e-5$), to assess the impact of batch size and learning rate in the pix2pix method
- **Experiment (2):** Compare *PCA* (E) and *pix2pix* ($E, 1, 5e-5$), to assess the relative performance of the PCA-based and pix2pix methods
- **Experiment (3):** Compare *PCA* (E) and *pix2pix* ($E, 1, 5e-5$) to *PCA* (G) and *pix2pix* ($G, 1, 5e-5$), to assess the impact of the chosen gravity field functional

The reasoning behind those experiments is to first make a choice on the hyperparameters used for the pix2pix method, which is then kept for the two subsequent experiments, where the performance of the PCA-based and pix2pix signal separation methods is compared. The latter is done considering the prediction errors on the (complete, 1-year, global) test dataset as a quality measure. The prediction errors are computed as difference between the predicted and true

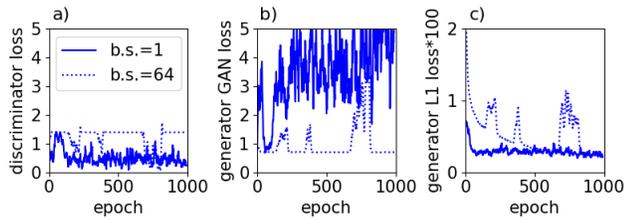


Figure 4: Loss curves for the two signal separation models $\text{pix2pix}(E,1,5e-5)$ (solid) and $\text{pix2pix}(E,64,5e-5)$ (dotted) as a function of training epoch. The two separation models only differ by their batch size. The learning rate amounts to $5e-5$ for both models, both models have been trained on EWH model data. The curves have been smoothed with a window of length 10 to improve the readability of the plots.

AO and H signals, which are given as 3d latitude-longitude-time matrices, as detailed in section 3.

For the visualizations, statistical measures such as the mean absolute value of all matrix entries, or root-mean-square (rms) values along two of the three matrix axes are built, or the values corresponding to a fixed point in time or a fixed point in space are extracted. Besides the individual AO and H prediction errors, also their sum is considered, which allows to classify the error type: If the component errors mostly cancel when summed up, this indicates that the prediction errors are dominated by parts of the input signal that are assigned to the wrong component (called "assignment error", in the following). If, in contrast, the component errors mostly add up as they have the same sign, the prediction errors are dominated by parts of the input signal that are not assigned to any of the two components (called "absent signal error", in the following). Besides the absolute prediction errors, also their magnitude relative to the signal amplitude is considered, by dividing the mean absolute error by the mean absolute value of the corresponding (AO, H or AOH) signal. In all cases, we treat the AO and H predictions of the PCA-based and the pix2pix separation models equally.

In addition, the training process of the pix2pix models is examined by computing the prediction errors on the test or training dataset after each training epoch. Also, internal quality measures of the pix2pix cGAN, being the discriminator loss as well as the two components of the generator loss (GAN loss and L1 loss) are investigated as a function of training epoch.

4.2 Experiment (1): Influence of hyperparameter choice in pix2pix cGAN

The purpose of this first experiment is to assess the impact of the hyperparameters batch size (b.s.) and learning rate (l.r.) on the performance of the pix2pix signal separation method. To this end, we consider the separation models $\text{pix2pix}(E,1,5e-5)$, $\text{pix2pix}(E,64,5e-5)$, $\text{pix2pix}(E,1,1e-4)$ and $\text{pix2pix}(E,1,1e-5)$ (cf. section 4.1) and evaluate internal (loss curves) as well as external quality measures (prediction error on test year 2002 data).

Firstly, we investigate the impact of the batch size by fixing the value of the learning rate to $5e-5$ and comparing the separation

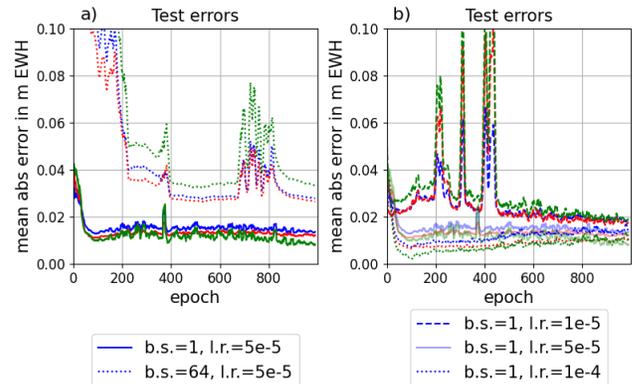


Figure 5: Mean absolute prediction errors on the test dataset (year 2002), as a function of training epoch, for the signal separation models $\text{pix2pix}(E,1,5e-5)$ and $\text{pix2pix}(E,64,5e-5)$ (a) and $\text{pix2pix}(E,1,1e-5)$, $\text{pix2pix}(E,1,5e-5)$ and $\text{pix2pix}(E,1,1e-4)$ (b). Blue (red) curves show the AO (H) prediction errors; green curves show the errors on the sum of the predicted AO and H components. The curves have been smoothed with a window of length 10 to improve the readability of the plots.

models $\text{pix2pix}(E,1,5e-5)$ and $\text{pix2pix}(E,64,5e-5)$. Fixing the learning rate to different values revealed similar results (not included here).

As shown by Fig. 4, in the case of $\text{pix2pix}(E,1,5e-5)$, the discriminator and generator GAN losses fluctuate strongly as a function of epoch and do not seem to converge to stable values, while they stay (besides temporary deviations) at a relatively stable level for $\text{pix2pix}(E,64,5e-5)$. Another effect is that both the L1 loss (Fig. 4 c) and the prediction errors on the test dataset (Fig. 5 a) decrease more slowly for $\text{pix2pix}(E,64,5e-5)$ than for $\text{pix2pix}(E,1,5e-5)$, i.e. the training as a function of epoch is slowed down if increasing the batch size.

That is, for our specific application of the pix2pix cGAN architecture, $\text{b.s.} > 1$ is required to achieve a stable training of the GAN in terms of a simultaneous training of the discriminator and generator part. However, after 1000 training epochs, using $\text{b.s.}=1$ gives about half the prediction errors on the test dataset than $\text{b.s.}=64$. As our application aims at finding a separation model that gives minimum prediction errors on the test dataset in terms of the numerical difference between predicted and true signal components, we choose $\text{b.s.}=1$ for our signal separation application of pix2pix .

The impact of the learning rate on the test errors is visualized by Fig. 5 b, where $\text{pix2pix}(E,1,1e-5)$, $\text{pix2pix}(E,1,5e-5)$ and $\text{pix2pix}(E,1,1e-4)$ are compared. $\text{pix2pix}(E,1,1e-4)$ shows a strong decrease of errors at the beginning of the training and a slight increase starting at about epoch 100, indicating a too long training. In the case of $\text{pix2pix}(E,1,1e-5)$, the smaller learning rate leads to large spikes in the test error curves indicating local minima that are disturbing the training process. Also, the model does not seem to have converged after 1000 epochs yet. That is, changing the learning rate in the pix2pix method mainly impacts the regularity of the training process and the optimum number of epochs to be used to reach the best possible performance on the test dataset.

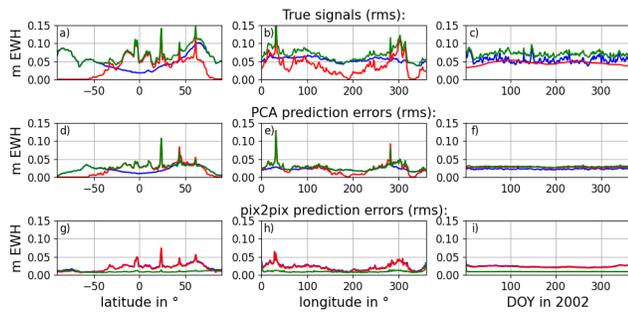


Figure 6: Prediction errors of the signal separation models $PCA(E)$ (d-f) and $pix2pix(E, 1.5e-5)$ (g-i) on the test dataset (year 2002) in terms of RMS values built along the time and longitude axes (left), the time and latitude axes (middle) and the latitude and longitude axes (right). The blue (red) curves refer to the AO (H) signal component, while the green curve refers to their sum.

Considering the above-made findings, we use $b.s.=1$ and $l.r.=5e-5$ for demonstrating the performance of the $pix2pix$ cGAN architecture for the considered signal separation task in the following two experiments (sections 4.3 and 4.4). An optimization of the values for $b.s.$, $l.r.$ and the training duration for our setting would require further systematic tests.

4.3 Experiment (2): Relative performance of PCA and $pix2pix$ cGAN

The purpose of this second experiment is to assess the relative performance of the $pix2pix$ and the PCA -based method for our signal separation task. To this end, we compute the prediction errors of the separation models $pix2pix(E, 1.5e-5)$ and $PCA(E)$ on the complete test dataset, and visualize their rms values along the latitude, longitude and time axes in Fig. 6.

It can be seen that the prediction errors of both methods are mostly smaller than the corresponding signals, on the order of 50 % of the signal amplitudes. A systematic difference between the two methods is their behavior regarding the sum of the component errors: Following our explanations in section 4.1, in the case of $pix2pix(E, 1.5e-5)$, the prediction errors seem to be dominated by "assignment errors", while in the case of $PCA(E)$, "absent signal errors" seem to dominate.

The former observation indicates that the $pix2pix(E, 1.5e-5)$ neural net has learned during the training to distribute (almost) the total AOH signal energy to the two available components, despite the fact that no constraint forcing the sum of the predicted components being equal to the input signal was imposed.

The latter observation indicates that in the $PCA(E)$ separation model, the AO and H subspaces representing the AO and H signal variability seem to be mostly distinct, having a negligible overlap, i.e. $PCA(E)$ seems to be able to distinguish well between signal parts caused by AO and H processes. The amplitude of the "absent signal errors", however, indicates that the dimensionality of the two component subspaces is not sufficient to resolve the full AO and H signal variability in the test dataset, such that a significant part

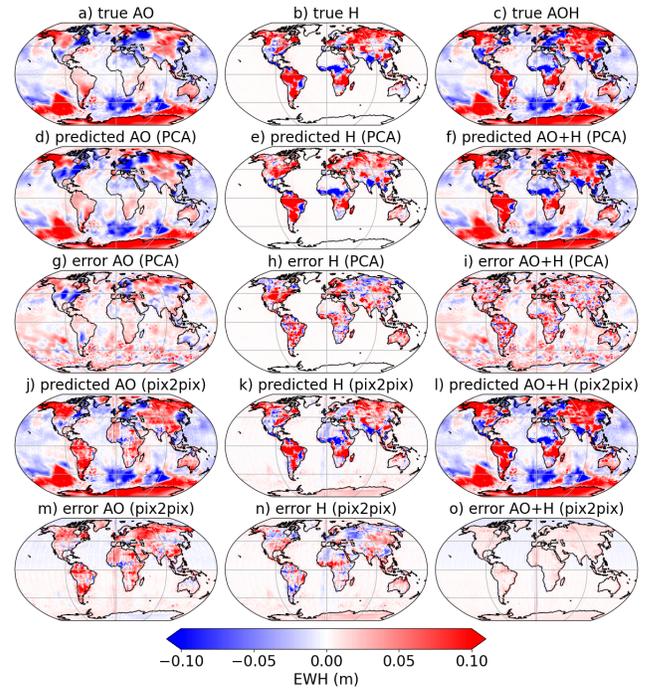


Figure 7: Predictive performance of the signal separation models $PCA(E)$ (d-i) and $pix2pix(E, 1.5e-5)$ (j-o) on the test dataset (year 2002). The data of the time step 100 has been extracted to observe the respective spatial patterns.

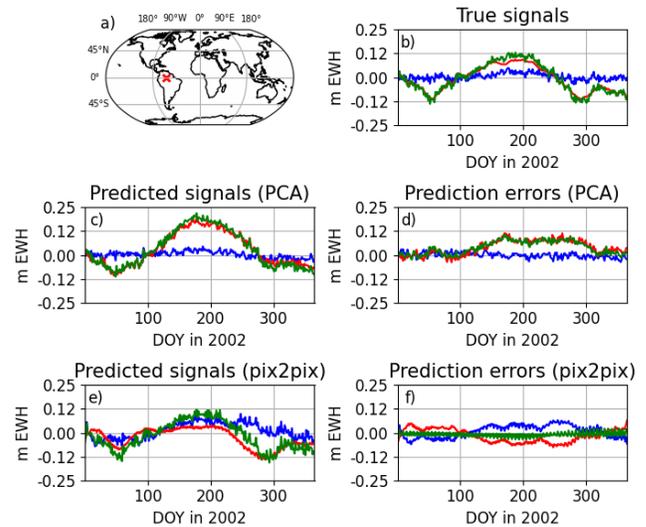


Figure 8: Predictive performance of the signal separation models $PCA(E)$ (c, d) and $pix2pix(E, 1.5e-5)$ (e, f) on the test dataset (year 2002). The data of a coordinate grid cell in the Amazon basin has been extracted to observe the respective temporal patterns.

of the input signal can not be assigned to any of the components. This could be improved by using a larger number of EOFs to span the individual subspaces, which would however require to use a larger amount of data (e.g. including the data of multiple years) for the computation of the EOFs.

Interestingly, for both methods, the type of prior knowledge given to the respective method is not (clearly) reflected by the prediction error variations along the individual coordinate directions: Although in *PCA* (*E*) only the signal patterns in space (latitude, longitude) domain are used as prior knowledge, also the variability in time domain is reproduced by the model. In the case of *pix2pix* (*E*, $1.5e-5$), despite only the signal patterns along the longitude and time axes were explicitly given to the model in the training process, and signals of different latitudes are treated as independent, the model does not perform worse in reproducing the signal variability along the latitude compared to the longitude axis.

Figure 7 shows the signals and errors for one point in time, the choice of which is arbitrary and does not impact the qualitative spatial behavior observed (which is also reflected by the very weak time-dependency of the prediction errors, revealed by Fig. 6 f and i). Figure 7 shows that the AO and H predictions of *PCA* (*E*) and *pix2pix* (*E*, $1.5e-5$) agree well in their large-scale pattern with the true signals, and their errors are dominated by short-scale features. Besides the observation of (mostly) cancelling AO and H prediction errors in the case of *pix2pix* (*E*, $1.5e-5$) and (mostly) additive component errors in the case of *PCA* (*E*) already made above, a closer look to the *PCA* (*E*) results allows to give examples of the two prediction error types defined in section 4.1: "Absent signal errors" can be observed in Fig. 7 h in South America and Africa, where the H errors consist of short-scale H signal energy, as the AO signals in these regions are rather low. "Assignment errors" can be observed in the more large-scale AO and H error patterns on the continents in the northern hemisphere, which mostly cancel when added up (cf. Fig. 7 i).

Furthermore, Fig. 7 m and n reveal that the "assignment errors" of *pix2pix* (*E*, $1.5e-5$) are mostly confined to the continents, i.e. the neural net learned that AOH signal energy in oceanic regions has to be assigned to the AO component, as the H signal is zero there.

Figure 8 shows the signals and errors for a fixed continental grid point with dominating H signal, which has been chosen to test the separation models' performance to extract the H signal in the Amazon basin. The shown temporal evolution of the signals and errors confirms the observations made in Fig. 7 on the error types in this region: In the case of *PCA* (*E*), the AO and H errors mostly add up, while they nearly cancel in the case of *pix2pix* (*E*, $1.5e-5$).

In summary, we found that the *pix2pix* method provides prediction errors of comparable amplitude as the *PCA* method. Remarkably, the *pix2pix* neural net was even able to learn relationships such as the equality between the sum of the predicted components and the input signal, or the confinement of the H component to the continents, without having received explicit information on these rules via corresponding constraints. We found that in terms of the predictive performance on the test dataset, the model *pix2pix* (*E*, $1.5e-5$) is limited by "assignment errors", which are signal parts that are assigned to the wrong signal component, while the model *PCA* (*E*) is limited by an insufficient representation of especially short-spatial scale signal components by the used AO and H subspaces, leading to a dominance of "absent signal errors".

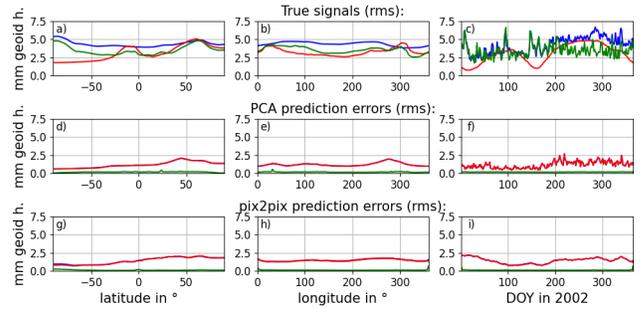


Figure 9: Prediction errors of the signal separation models *PCA* (*G*) (d-f) and *pix2pix* (*G*, $1.5e-5$) (g-i) on the test dataset (year 2002) in terms of RMS values built along the time and longitude axes (left), the time and latitude axes (middle) and the latitude and longitude axes (right). The blue (red) curves refer to the AO (H) signal component, while the green curve refers to their sum

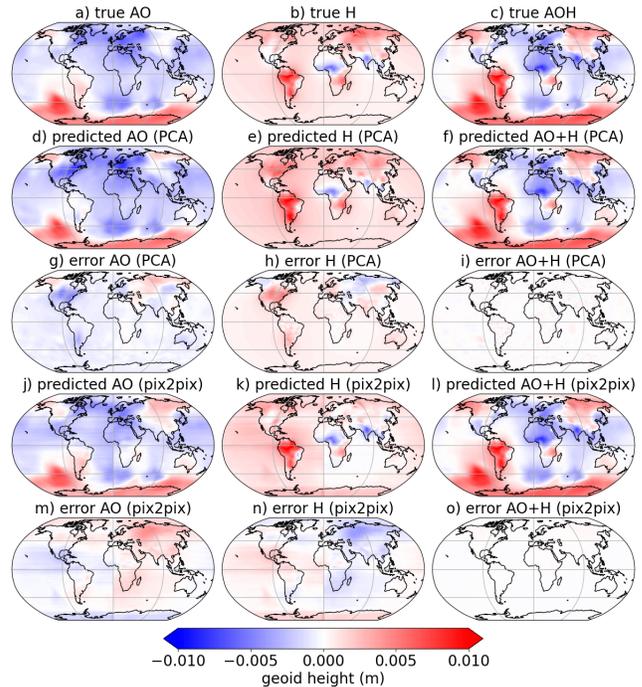


Figure 10: Predictive performance of the signal separation models *PCA* (*G*) (d-i) and *pix2pix* (*G*, $1.5e-5$) (j-o) on the test dataset (year 2002). The data of the time step 100 has been extracted to observe the respective spatial patterns.

4.4 Experiment (3): Influence of used gravity field functional

In this section, we investigate how the results presented in section 4.3 change if signals preprocessed to data grids of geoid heights instead of EWHs are considered (cf. section 2). To this end, we investigate the predictive performance of the separation models

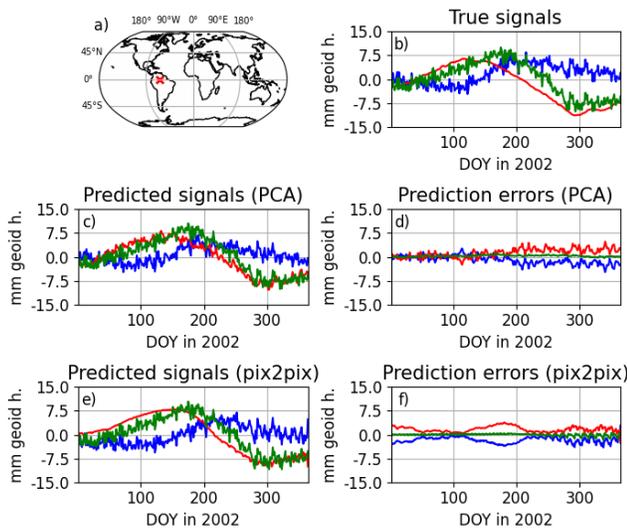


Figure 11: Predictive performance of the signal separation models *PCA* (*G*) (c, d) and *pix2pix* (*G*, $1.5e-5$) (e, f) on the test dataset (year 2002). The data of a coordinate grid cell in the Amazon basin has been extracted to observe the respective temporal patterns.

PCA (*G*) and *pix2pix* (*G*, $1.5e-5$) in Figs. 9, 10 and 11, and compare them to the before-analyzed Figs. 6, 7 and 8 that show corresponding results for the separation models *PCA* (*E*) and *pix2pix* (*E*, $1.5e-5$). Additionally, the prediction errors of *PCA* (*E*), *PCA* (*G*), *pix2pix* (*E*, $1.5e-5$) and *pix2pix* (*G*, $1.5e-5$) on the training and test datasets are compared in Fig. 12. We note that the below-presented results on the impact of using geoid height vs. EWH grid data have been found to be independent of the batch size and learning rate used in the *pix2pix* method.

Regarding the considered signals themselves, Fig. 9 a-c and 10 a-c reveal that, when expanded to geoid height grids, the signals have a much more long-wavelength signature in space domain compared to the EWH grid data (cf. Fig. 6 a-c and 7 a-c). This is due to the amplification of the small-scale signal components when expanding the SH signal coefficients to EWH grids (cf. section 2).

Regarding the prediction errors of *PCA* (*G*) and *pix2pix* (*G*, $1.5e-5$), Fig. 9 shows that, similarly to what was observed for *PCA* (*E*) and *pix2pix* (*E*, $1.5e-5$), both methods predict the AO and H signals at a comparable accuracy, with a signal-to-noise ratio of > 1 .

One difference to the results presented in section 4.3 is the predictive performance of the *PCA*-based and the *pix2pix* method along the time axis: As shown by Figs. 9 f, i and 11 d, f, the *PCA* (*G*) prediction errors fluctuate more strongly in time domain than it is the case for *pix2pix* (*E*, $1.5e-5$). Also, Fig. 11 reveals that the *pix2pix* method seems to be better at distinguishing the low-frequency, annual behavior of the H component from the higher-frequency behavior of the AO signals. A reason could be that the *PCA*-based method only uses prior knowledge in terms of spatial patterns while the *pix2pix* method is also trained on temporal patterns.

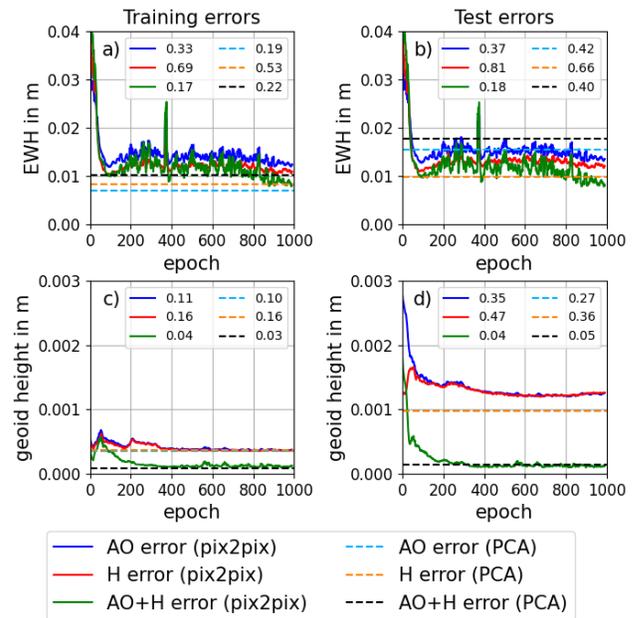


Figure 12: Mean absolute prediction errors on the training (year 1995; a, c) and test (year 2002; b, d) datasets, as a function of training epoch, for the separation models *pix2pix* (*E*, $1.5e-5$) (a, b) and *pix2pix* (*G*, $1.5e-5$) (c, d). Dashed lines show the respective error of the *PCA* (*E*) (a, b) and *PCA* (*G*) (c, d) predictions for reference. The numbers in the insets give the mean absolute error relative to the respective mean absolute signal as achieved by the separation models (after 1000 training epochs, in the case of the *pix2pix* models).

A systematic difference of *PCA* (*E*) and *PCA* (*G*), as can be seen by comparing Figs. 6 d-f and 9 d-f (and also Figs. 7 g-i and 10 g-i), is that while the *PCA* (*E*) prediction errors are dominated by "absent signal errors", the *PCA* (*G*) prediction errors are dominated by "assignment errors". This means, if signals expanded to the more long-wavelength geoid height grids are investigated, also the *PCA* method distributes almost the total input signal to the two signal components AO and H, as it is the case for both *pix2pix* (*E*, $1.5e-5$) and *pix2pix* (*G*, $1.5e-5$).

The fact that the "absent signal errors" are smaller for *PCA* (*G*) than for *PCA* (*E*) indicates that in the case of *PCA* (*G*), the AO and H signal subspaces better represent the AO and H signal variability contained in the test dataset. This reduces the amount of input signal energy that does not become assigned to any of the two components. However, the fact that the "assignment errors" are larger for *PCA* (*G*) than for *PCA* (*E*) indicates larger correlations between the AO and H signals when expressed as geoid heights, which lead to an increased overlap of the AO and H signal subspaces. This overlap could be reduced by using a smaller number of EOFs to span the AO and H subspaces for our *PCA*-based signal separation method (which would, however, increase the "absent signal error", i.e. an optimum number of modes to reach minimum prediction errors would need to be determined).

Finally, we consider how the training process of the separation models $\text{pix2pix}(E, 1, 5e-5)$ and $\text{pix2pix}(G, 1, 5e-5)$ compares, by considering the associated training and test errors as a function of epoch in Fig. 12. As shown by panels a and b, the training and test error curves for $\text{pix2pix}(E, 1, 5e-5)$ look very similar and reach a comparable error level, while for $\text{pix2pix}(G, 1, 5e-5)$ (panels c and d), the AO and H test errors are about three times larger than the associated training errors. This discrepancy of the performance of $\text{pix2pix}(G, 1, 5e-5)$ on the training and test datasets suggests a strong overfitting when training based on geoid height grids, which seems to be avoided when training based on EWH grids. This can be explained by the geoid height grids being dominated by long-wavelength features, which lead to strong correlations among the training samples used for the pix2pix method. In the case of the EWH data samples, as here the smaller-scale structures dominate, these correlations are much reduced, preventing the separation model to "remember" signal structures instead of learning general rules on how to predict the target images based on the input images.

The prediction errors of the $\text{PCA}(G)$ and $\text{PCA}(E)$ separation models on the training and test datasets are indicated as dashed lines in Fig. 12. For both the $\text{PCA}(G)$ and the $\text{PCA}(E)$ model, the AO and H prediction errors on the test dataset are larger than on the training dataset. This reflects the fact that the EOFs spanning the AO and H signal subspaces were computed based on the training data, thereby representing the signal characteristics of the training data, while the test data's signal variability is not fully included in the used AO and H signal subspaces. While this is a general observation independent of the considered functional, the magnitude of the difference between training and test errors reflects how much the training and test data differ in terms of their spatial signal patterns.

The black dashed lines in Fig. 12 show that both for the training and the test errors, the AO and H prediction errors of $\text{PCA}(E)$ accumulate when added up, while they mostly cancel in the case of $\text{PCA}(G)$. This confirms the above-made observation on dominating "absent signal errors" in the case of $\text{PCA}(E)$ and dominating "assignment errors" in the case of $\text{PCA}(G)$.

The relative errors given by the insets in Fig. 12 allow to judge the mean absolute errors in relation to the corresponding signals. As the mean absolute value of the AO EWH signal is 2.5-times (1.3-times, for geoid heights) larger than the respective H signal, the relative AO errors are smaller than the relative H errors if the same absolute prediction error is reached.

A comparison of the two investigated methods based on their performance on the test dataset in Fig. 12 reveals that when working with geoid heights, the relative errors for AO and H achieved by $\text{PCA}(G)$ are about 20 % smaller than the associated $\text{pix2pix}(G, 1, 5e-5)$ errors. This indicates that here, the overfitting problem of the pix2pix model seems to be larger than the problem due to the overlapping AO and H signal subspaces in the case of the PCA -based method.

When working with EWH data, a choice between the two methods seems less obvious, since the relative AO error of $\text{pix2pix}(E, 1, 5e-5)$ is 12 % smaller than the corresponding $\text{PCA}(E)$ error, but the relative H error of $\text{PCA}(E)$ is 19 % smaller than the corresponding $\text{pix2pix}(E, 1, 5e-5)$ error. However, the $\text{pix2pix}(E, 1, 5e-5)$ error curves do not seem to have converged after 1000 epochs yet, i.e. a longer training could lead to a further improvement of the pix2pix model.

5 CONCLUSIONS AND OUTLOOK

In this paper, we investigate the performance of two signal separation methods, one based on PCA (cf. section 3.1) and one based on the pix2pix cGAN neural network architecture (cf. section 3.2), on the task of separating atmosphere/ocean and hydrological spatial-temporal gravity signals from their sum. Both methods were shown to predict the two signal components at a signal-to-noise ratio of > 1 , at an error level between 27 and 81 % of the respective signal, depending on the specific separation model.

In general, two types of errors could be identified to contribute to the prediction errors in the considered signal separation task:

- (1) parts of the input signal that are assigned to the wrong signal component, which we call "assignment errors"
- (2) parts of the input signal that are not assigned to any of the two signal components, which we call "absent signal errors"

In the case of the pix2pix method, the prediction errors are dominated by the first error type, indicating that the total input signal energy becomes distributed to the two signal components. This was learned solely based on the training data, as no corresponding constraint on the sum of the predicted components has been applied.

In the case of the PCA method, which of the two error types dominates the prediction errors depends on the relationship of the signal subspaces used to represent the signal variability of the two signal components: If the subspaces show a considerable overlap, which is the case if the two signal components are spatially highly correlated, the first error type dominates. The opposite is true if the overlap is negligible. In the framework of our analysis, the first situation has been observed if gravity signals expanded in geoid height grids were considered, while the second situation was present when working with EWH grids.

The choice of the gravity field functional to which the SH gravity coefficients are expanded also has a major impact on the pix2pix method: As they are dominated by the large-scale signal components, using geoid height grids introduces significant correlations among the pix2pix training samples, leading to a strong overfitting effect. This effect is resolved by using EWH data grids which are more strongly dominated by the short-scale signal structures.

In general, we find the very different behavior of both separation methods when applied to either EWH or geoid height data noteworthy, because both quantities have been derived from the same input model data (ESA ESM). Therefore, from a theoretical point of view EWH and geoid height contain the same information, but only the spectral components are differently weighted.

For the further development of the two signal separation methods, we recommend the expansion of the SH gravity data to EWH grids, as this has the advantage to both avoid the overfitting problem in the case of the pix2pix method, as well as the problem of correlated AO and H EOFs in the case of the PCA method. To improve the predictive ability of the algorithms, we suggest to include more data in the model building stage, which in the case of the PCA method would increase the dimensionality of the signal subspaces, such that they become representative of a larger possible variability of the two signal components. Also in the pix2pix method, an augmentation of the training dataset is expected to further reduce the prediction errors.

In order to develop the algorithms towards signal separation in real satellite gravity data, the considered signal separation task needs to be extended to include a larger number of individual gravity signal components to be separated, including a noise component. In this context, we suggest to subtract geophysical model predictions for the larger-amplitude signal components before the separation step, in which the (residual) signal components are then estimated. This would especially reduce the "assignment errors" of smaller-amplitude components relative to their signal amplitude.

Regarding the pix2pix neural network method, section 4.2 revealed that a hyperparameter configuration that provides a good signal separation performance on the test dataset does not necessarily also provide an equilibrated state between the discriminator and the generator part of the GAN architecture. In particular, a divergence of the GAN losses with decreasing values for the discriminator and increasing values for the generator was observed. For the considered task, the numerical difference between the generated and true output is the primary performance measure, while the visual plausibility of the generated output as it is judged by the discriminator is secondary. This raises the question whether the optimization of the GAN loss terms regulating the balance between discriminator and generator contributes at all to the signal separation performance of the considered networks. Therefore, we recommend to test simpler architectures such as U-Nets for our task in future studies.

Further questions to be addressed in the future regarding a neural network-based signal separation method comprise investigating the impact of various sampling strategies besides the slicing of the data at constant latitudes as done in the present study, assessing the role of individual signal features such as specific frequencies or spatial patterns in the training process, as well as specifically designing the algorithm to address the characteristics of the considered spatial-temporal dataset, involving the introduction of more prior knowledge to the system in the form of additional loss terms.

ACKNOWLEDGMENTS

Funding was provided by the German Research Foundation (DFG) research training group UPLIFT 440512084 (GRK 2698). We also acknowledge the provision of supercomputing resources by the Leibniz Supercomputing Centre (LRZ; Address: Boltzmannstraße 1, 85748 Garching bei München, Germany).

REFERENCES

- [1] Eva Boergens, Elena Rangelova, Michael G. Sideris, and Juergen Kusche. 2014. Assessment of the capabilities of the temporal and spatiotemporal ICA method for geophysical signal separation in GRACE data. *Journal of Geophysical Research: Solid Earth* 119, 5 (2014), 4429–4447. <https://doi.org/10.1002/2013JB010452>
- [2] Jianli Chen, Anny Cazenave, Christoph Dahle, William Llovel, Isabelle Panet, Julia Pfeffer, and Lorena Moreira. 2022. Applications and Challenges of GRACE and GRACE Follow-On Satellite Gravimetry. *Surveys in Geophysics* 43, 1 (2022), 305–345. <https://doi.org/10.1007/s10712-021-09685-x>
- [3] Henryk Dobslaw, Inga Bergmann-Wolf, Robert Dill, Ehsan Forootan, Volker Klemann, Jürgen Kusche, and Ingo Sasgen. 2015. The updated ESA Earth System Model for future gravity mission simulation studies. *Journal of Geodesy* 89, 5 (2015), 505–513. <https://doi.org/10.1007/s00190-014-0787-8>
- [4] Frank Flechtner, Frank Webb, Michael Watkins, Felix Landerer, Christoph Dahle, and Srinivas Bettadpur. 2017. Current status of the GRACE follow-on mission. *Geophysical Research Abstracts* 19, EGU 2017–4566 (2017).
- [5] E. Forootan and J. Kusche. 2011. Separation of global time-variable gravity signals into maximally independent components. *Journal of Geodesy* 86, 7 (2011), 477–497. <https://doi.org/10.1007/s00190-011-0532-5>
- [6] E. Forootan, R. Rietbroek, J. Kusche, M.A. Sharifi, J.L. Awange, M. Schmidt, P. Omond, and J. Famiglietti. 2014. Separation of large scale water storage patterns over Iran using GRACE, altimetry and hydrological data. *Remote Sensing of Environment* 140 (2014), 580–595. <https://doi.org/10.1016/j.rse.2013.09.025>
- [7] Betty Heller-Kaikov, Roland Pail, and Ilias Daras. 2023. Mission design aspects for the mass change and geoscience international constellation (MAGIC). *Geophysical Journal International* 235, 1 (2023), 718–735. <https://doi.org/10.1093/gji/ggad266>
- [8] B. Hofmann-Wellenhof and Helmut Moritz. 2005. *Physical Geodesy*. Springer, 403 pages.
- [9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2016. Image-to-Image Translation with Conditional Adversarial Networks. (2016). <https://doi.org/10.48550/arXiv.1611.07004>
- [10] Richard P. Kornfeld, Bradford W. Arnold, Michael A. Gross, Neil T. Dahya, William M. Klipstein, Peter F. Gath, and Srinivas Bettadpur. 2019. GRACE-FO: The Gravity Recovery and Climate Experiment Follow-On Mission. *Journal of Spacecraft and Rockets* 56, 3 (2019), 931–951. <https://doi.org/10.2514/1.A34326>
- [11] Luca Massotti, Christian Siemes, Günther March, Roger Haagmans, and Pierluigi Silvestrin. 2021. Next Generation Gravity Mission Elements of the Mass Change and Geoscience International Constellation: From Orbit Selection to Instrument and Mission Design. *Remote Sensing* 13, 19 (2021), 3935. <https://doi.org/10.3390/rs13193935>
- [12] Gabriel Meseguer-Brocal and Geoffroy Peeters. 2019. Conditioned-U-Net: Introducing a Control Mechanism in the U-Net for Multiple Source Separations. (2019). <https://doi.org/10.48550/ARXIV.1907.01277>
- [13] Rudolph W. Preisendorfer and Curtis D. Mobley. 1988. Principal Component Analysis in Meteorology and Oceanography. <https://api.semanticscholar.org/CorpusID:122456032>
- [14] Matthew Rodell and James S. Famiglietti. 1999. Detectability of variations in continental water storage from satellite observations of the time dependent gravity field. *Water Resources Research* 35, 9 (1999), 2705–2723. <https://doi.org/10.1029/1999WR900141>
- [15] Martin Schmeer, Michael Schmidt, Wolfgang Bosch, and Florian Seitz. 2012. Separation of mass signals within GRACE monthly gravity field models by means of empirical orthogonal functions. *Journal of Geodynamics* 59-60 (2012), 124–132. <https://doi.org/10.1016/j.jog.2012.03.001>
- [16] Tianyan Shi, Yoichi Fukuda, Koichiro Doi, and Jun'ichi Okuno. 2022. Extraction of GRACE/GRACE-FO observed mass change patterns across Antarctica via independent component analysis (ICA). *Geophysical Journal International* 229, 3 (2022), 1914–1926. <https://doi.org/10.1093/gji/ggac033>
- [17] Xiao Sun, Jindong Xu, Yongli Ma, Tianyu Zhao, Shifeng Ou, and Lizhi Peng. 2020. Blind image separation based on attentional generative adversarial network. *Journal of Ambient Intelligence and Humanized Computing* 13, 3 (2020), 1397–1404. <https://doi.org/10.1007/s12652-020-02637-0>
- [18] B. D. Tapley, S. Bettadpur, M. Watkins, and C. Reigber. 2004. The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters* 31, L09607 (2004). <https://doi.org/10.1029/2004gl019920>
- [19] I. Velicogna and J. Wahr. 2013. Time-variable gravity observations of ice sheet mass balance: Precision and limitations of the GRACE satellite data. *Geophysical Research Letters* 40, 12 (2013), 3055–3063. <https://doi.org/10.1002/grl.50527>
- [20] Bramha Dutt Vishwakarma, Yann Ziegler, Jonathan L Bamber, and Sam Royston. 2022. Separating GIA signal from surface mass change using GPS and GRACE data. *Geophysical Journal International* 232, 1 (2022), 537–547. <https://doi.org/10.1093/gji/ggac336>
- [21] John Wahr, Mery Molenaar, and Frank Bryan. 1998. Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *Journal of Geophysical Research: Solid Earth* 103, B12 (1998), 30205–30229. <https://doi.org/10.1029/98jb02844>
- [22] DeLiang Wang and Jitong Chen. 2018. Supervised Speech Separation Based on Deep Learning: An Overview. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26, 10 (2018), 1702–1726. <https://doi.org/10.1109/TASLP.2018.2842159>
- [23] Carl Wunsch and Detlef Stammer. 1997. Atmospheric loading and the oceanic "inverted barometer" effect. *Reviews of Geophysics* 35, 1 (1997), 79–107. <https://doi.org/10.1029/96rg03037>
- [24] Yann Ziegler, Bramha Dutt Vishwakarma, Aoiheann Brady, Stephen Chuter, Sam Royston, Richard M Westaway, and Jonathan L Bamber. 2022. Can GPS and GRACE data be used to separate past and present-day surface loading in a data-driven approach? *Geophysical Journal International* 232, 2 (2022), 884–901. <https://doi.org/10.1093/gji/ggac365>