**Supplementary Information**

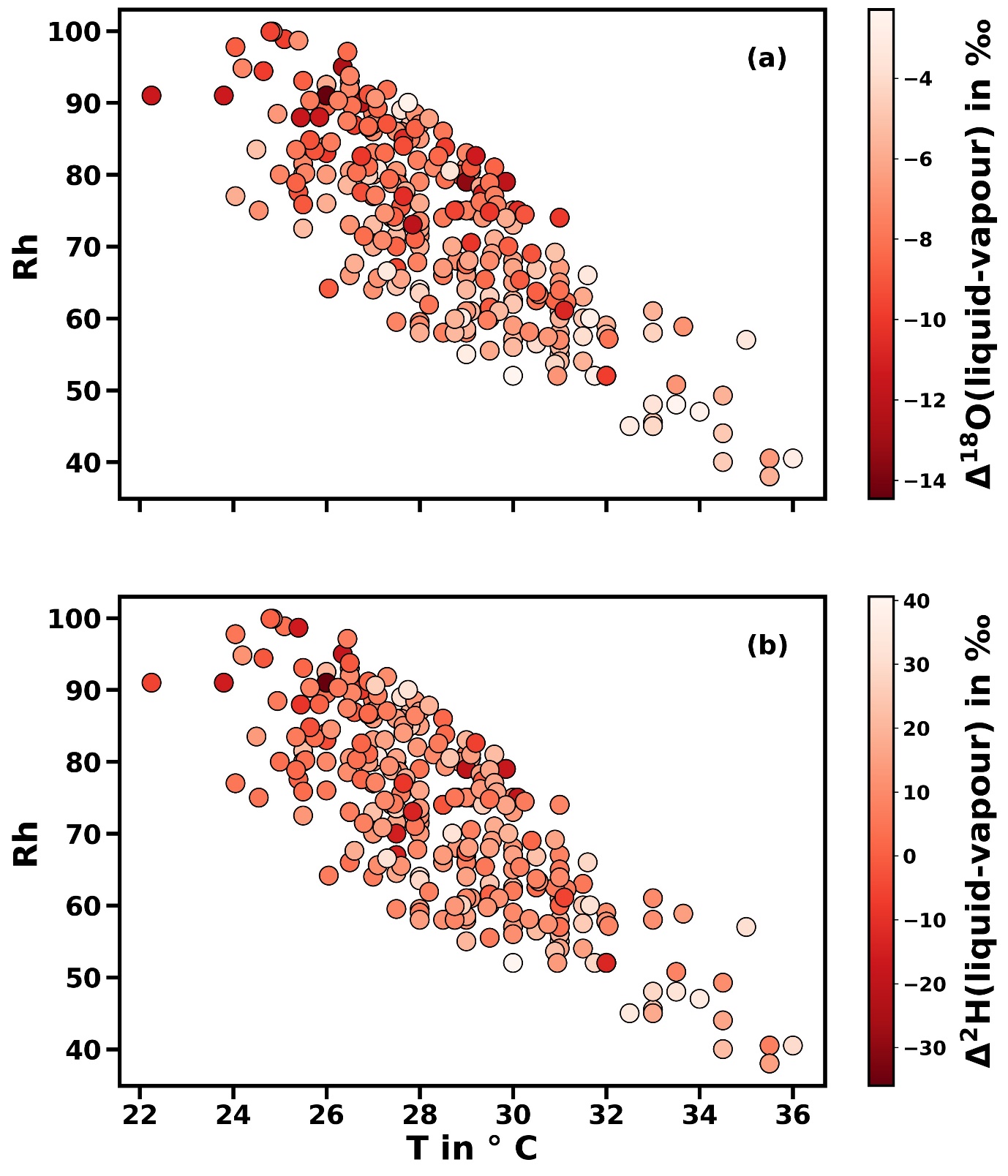
Akash Ganguly\*,[[1]](#footnote-1),2, Virendra Padhya1, Harsh Oza1, Gerhard Strauch3, R.D. Deshpande1

# 1Geosciences Division, Physical Research Laboratory, Navrangpura, Ahmedabad, 380009, India

# 2Indian Institute of Technology Gandhinagar, Gandhinagar, 382355, India

3Department of Hydrogeology, UFZ – Helmholtz Centre for Environmental Research, Germany

**Supplementary Material**

Figure S1. (a)- The isotopic difference ( = liquid - vapor) in δ18O between each pair of liquid and vapor samples is plotted against its T and RH, within a narrow range of Si (3.5 to 5.5). The darker shaded represents more depleted liquid compared to vapor, representing more fractionation. The greater extent of isotopic depletion due to more kinetic fractionation is observed under lower T - higher RH condition, within a narrow range of supersaturation; (b) The similar plot for δ2H.

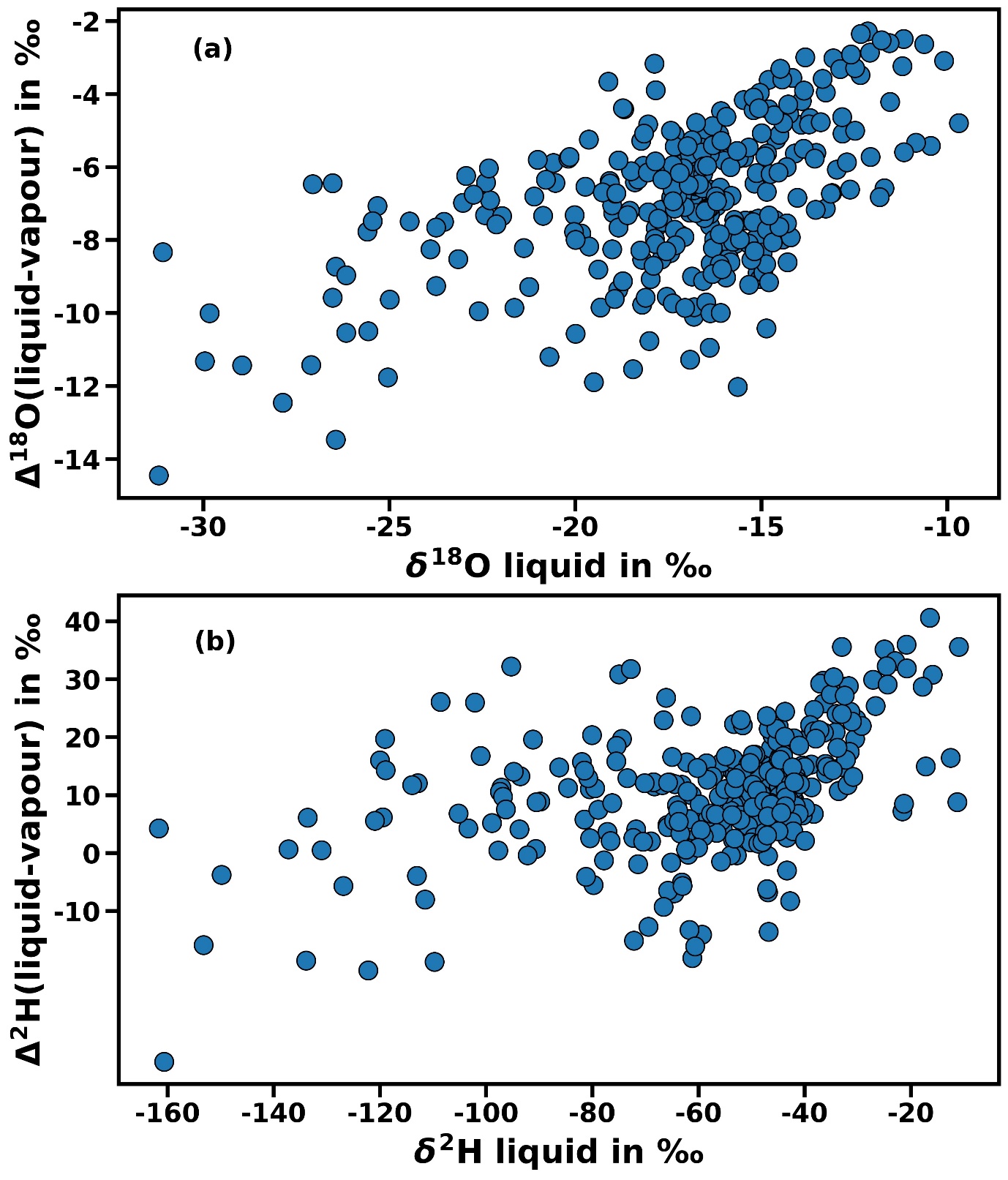


Figure S2. (a) The isotopic difference ( = liquid - vapor) in δ18O between each pair of liquid and vapor samples is plotted against δ18O of liquid, for a narrow range of Si (3.5 to 5.5) showing positive correlation between the two which confirms that isotopic depletion varies even under similar extent of supersaturation. (b) The similar plot for δ2H.

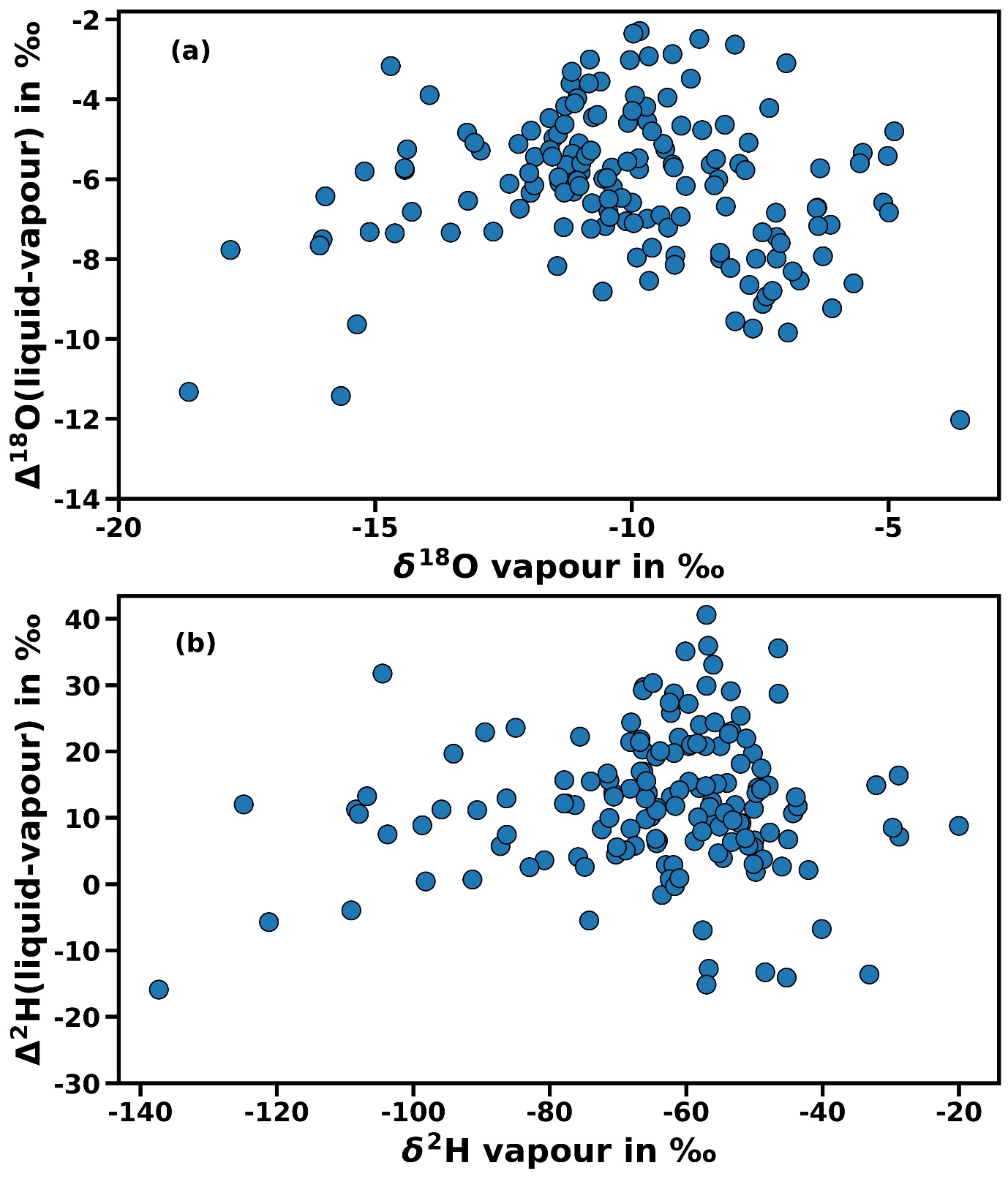
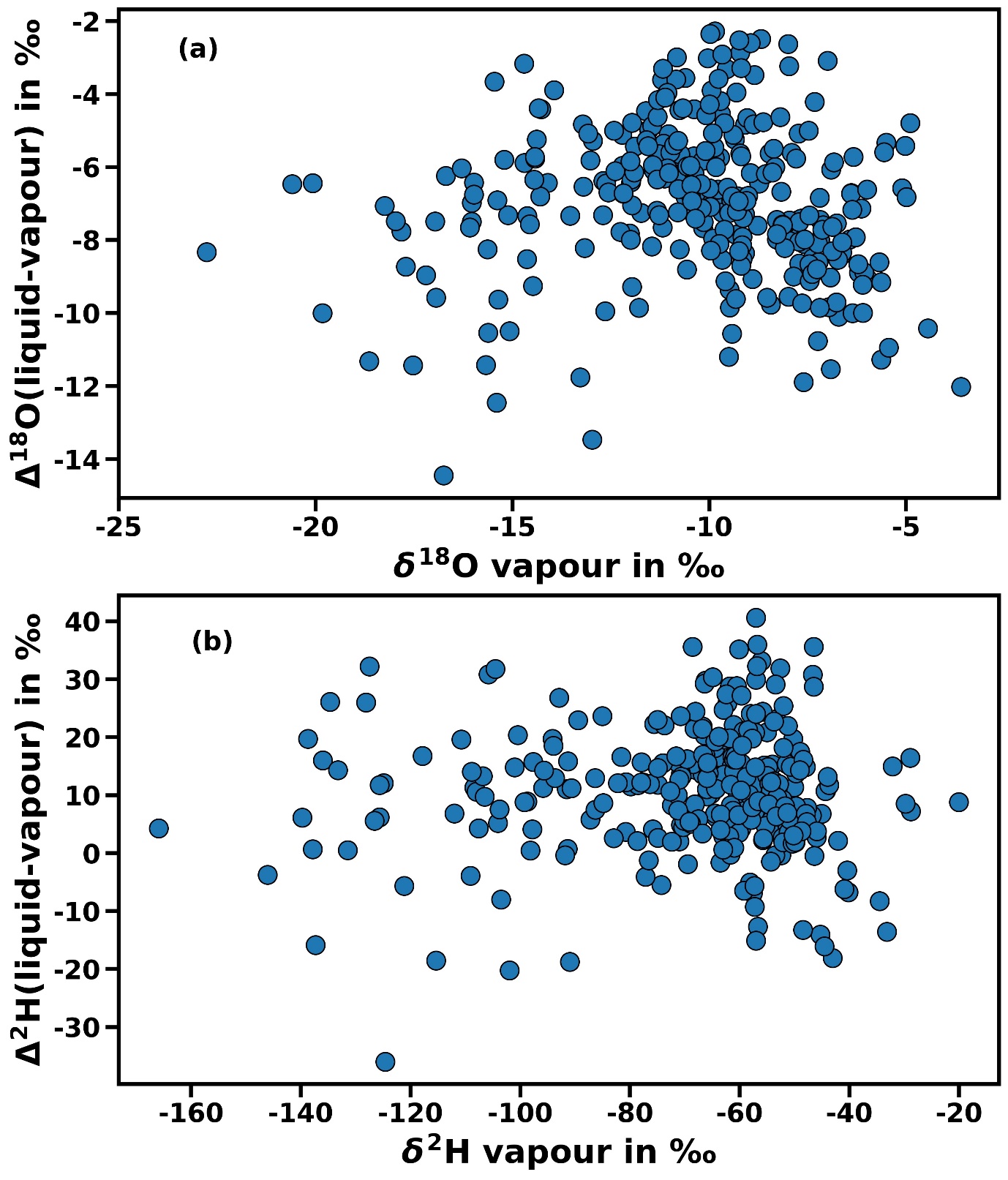


Figure S3. (a) The isotopic difference ( = liquid - vapor) in δ18O between each pair of liquid and vapor samples is plotted against δ18O of vapour for a narrow range of Si (3.5 to 4.5) showing no correlation between the two which confirms that isotopic depletion is independent of the δ18O of starting vapour. (b) The similar plot for δ2H.

 Figure S4. (a) The isotopic difference ( = liquid - vapor) in δ18O between each pair of liquid and vapor samples is plotted against δ18O of vapour for a narrow range of Si (3.5 to 5.5) showing no correlation between the two which confirms that isotopic depletion is independent of the δ18O of starting vapour. (b) The similar plot for δ2H.

|  |  |  |
| --- | --- | --- |
| **Model Architecture and**  **Performance** | **ANN for δ18O** | **ANN for δ**2H |
| Model layout | 5-10-10-1 | 5-20-18-16-15-1 |
| Activation layers | RELU (Rectified Linear Units) for hidden layers, Linear for final layer | RELU (Rectified Linear Units) for hidden layers, Linear for final layer |
| Batch size | 8 | 9 |
| Number of epochs | 120 | 125 |
| RMSE: On training dataset | 1.59 ‰ | 8.37 ‰ |
| RMSE: On test data set | 1.70 ‰ | 10.18 ‰ |
| Mean error on test data set | -0.24 ‰ | 0.53 ‰ |
| Standard deviation of test error | 1.69 ‰ | 10.19 ‰ |
| No of Train/Test Samples | 337,144 | 337,144 |
|  | | |

Table 1 ANN model architecture and performance in predicting vapour isotopic composition

The input layer dimensions (5 nodes) represents the number of feature vectors used for training while the final layer in case of a regression problem always has just 1 layer with a linear activation function. Since the non-linearity demonstrated for the oxygen isotope (18O) with respect to variation in diffusivity and other experimental limitations or constraints was less, a shallower neural network was chosen with 2 hidden layers having 10 nodes each and additional regularization was not deemed necessary. For the case of hydrogen isotopes, due to the intrinsic non-linearity in the problem statement owing to large range in Sic, a denser more complex network gave the best performance with 4 hidden layers having 20-18-16-15 nodes respectively.

We chose Rectified Linear units (RELU) as activation function after experimenting with RELU, sigmoid and tanh as possible choices. Since this is a relatively small dataset (~500 datapoints), we chose a relatively small batch size to update the hyper-parameters. Convergence was achieved for both models at around total ~120-125 iterations/ epochs after which there was not much improvement. Root mean squared error as well as mean absolute error were used as metrics for model validation whereas mean squared error was chosen as the loss function.

We employed Dropout (0.1), i.e. 10% of interlinkages/synapses between nodes were dropped for 2H in order to prevent overfitting whereas no regularisation was required for 18O. Further we used 5 fold cross validation in our training to ensure generalisation and robustness on an unknown dataset. The whole dataset was split randomly into 70-30 % with 70 % data for training the model and 30 % for model testing and validation. GridSearchCV was employed for hyperparameter tuning by running maximum permutations and combinations of different model architecture with different number of hidden layers and varying activation functions.

The relatively comparable values of both RMSE (root mean squared error) and mean of absolute error for both models hint at a robust fit without overfitting. The RMSE for δ18O is almost an order of magnitude lower than that for δ2H, but we must realise RMSE is not robust to outliers and is also dependent on the range of the target variable. For δ18O, we have a range of 19.16 ‰ for vapour, for which RMSE around 1.7 ‰ is quite reliable. For δ2H, we have a range in vapour of 147.73 ‰, for which an RMSE score of ≈ 10 ‰ is still quite reasonable. The tendency of mean error to lie near 0 further indicates the absence of a strong negative or positive bias in either case. Hence given a liquid sample, Rh and T, we can arrive at vapour isotope value within -0.24±1.69 ‰ for δ18O and 0.53±10.19 ‰ for δ2H with 1 standard deviation.

However, it must be noted that this uncertainty mentioned is only that of the ANN model and does not include the analytical uncertainty arising out of IRMS measurements. Hence, we need to account for the analytical uncertainty of 0.1 ‰ for δ18O and 1 ‰ for δ2H coming from IRMS. It is reasonable to assume the sources of both uncertainties are mutually exclusive, thus we can arrive at combined uncertainty for each tracer as

.

We can arrive at vapour isotope value within -0.24±1.79 ‰ for δ18O and 0.53±11.23 ‰ for δ2H, which provides a viable novel alternative way of sampling vapour isotopes over a large network simultaneously, especially in remote areas having logistical limitations.

1. \* Corresponding author- Akash Ganguly,

   Email – akashg@prl.res.in [↑](#footnote-ref-1)