Supporting Information

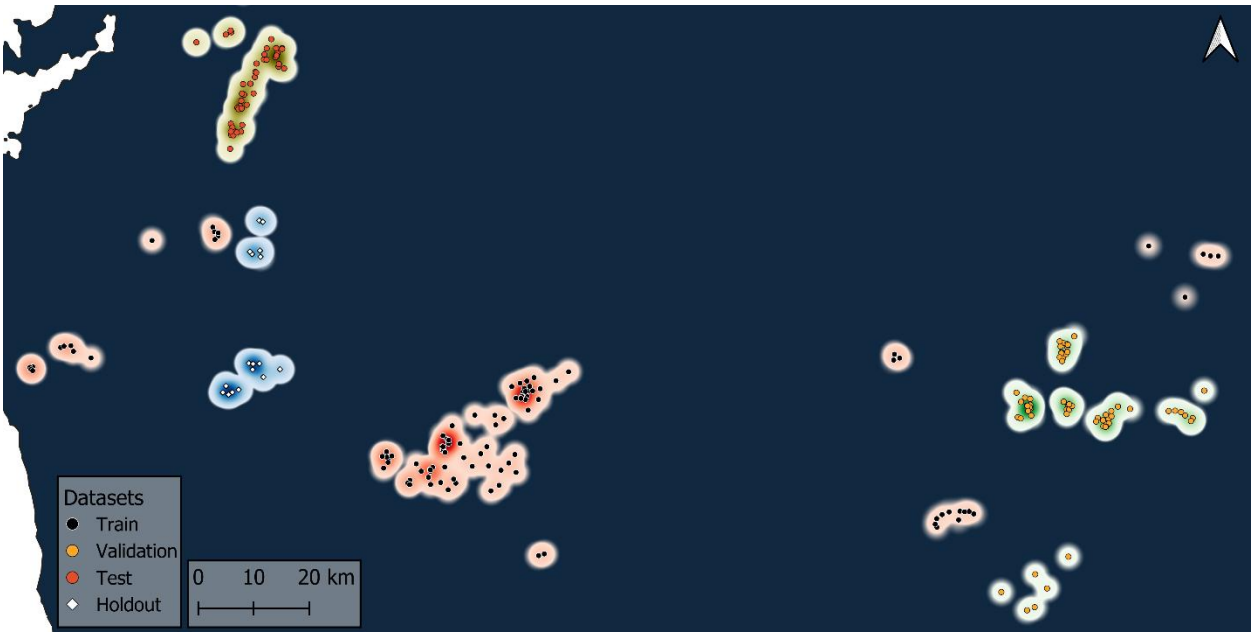

Figure 1 - Spatial distribution of sampling points by dataset, with heatmap (3km radius) illustrating the number of hours recorded. Heatmaps were generated with a 3km buffer.

Table 1 - HOWlish architecture. Layers in blue were frozen during the train process and therefore share weights with VGGish. "Classifier", "batch_normalization" and "activation" are the layers added by us.

Layer (type layer)	type	output shape	Parameters (n)
input_1	Input layer	96, 64, 1	0
conv1	2D convolution	96, 64, 64	640
pool1	Max pooling	48, 32, 64	0
conv2	2D convolution	48, 32, 128	73856
pool2	Max pooling	24, 16, 128	0
conv3/conv3_1	2D convolution	24, 16, 256	295168
conv3/conv3_2	2D convolution	24, 16, 256	590080
pool3	Max pooling	12, 8, 256	0
conv4/conv4_1	2D convolution	12, 8, 512	1180160
conv4/conv4_2	2D convolution	12, 8, 512	2359808
pool4	Max pooling	6, 4, 512	0
flatten	Flaten	12288	0
vggish_fc1/fc1_1	Fully connected	4096	50335744
vggish_fc1/fc1_2	Fully connected	4096	16781312
vggish_fc2	Fully connected	128	524416
classifier	Fully connected	1	129
batch_normalization	Batch normalization	1	4
activation	Sigmoid activation	1	0

Equations 1 - Equations for all measures of model performance reported, formulated with True Positives (TP), True Negatives (TN, False Positives (FP) and False Negatives (FN).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$False\ Positive\ Rate = \frac{FP}{FP + TN}$$

$$F_\beta\ score = (1 + \beta^2)\frac{(1 + \beta^2) \cdot TP}{(1 + \beta^2) \cdot TP + \beta^2 \cdot FN + FP}$$