Methods Section

## Model

We used a hierarchical framework to estimate Bayesian means of total emissions and resource use across five stressors (greenhouse gas [kg CO2-eq], nitrogen [kg N-eq], phosphorus [kg P-eq], land [m2a] and water [m3]). We do this for all unique scientific names (e.g., Cyprinidae, *Cyprinus carpio*) found in our compiled LCA data as well as for aggregated groups of taxa (e.g., miscellaneous carps). Specifically, our model structure has study , nested in scientific name , nested in taxa group . All Bayesian models were run using Markov Chain Monte Carlo (MCMC) algorithms in Stan (Stan Development Team, 2020) and called from R (R Core Team, 2020) using the R package, (Stan Development Team, 2019). All models converged (i.e., diagnostics for all parameters showed 0.99 < < 1.01, > 0.1, and no divergent transitions) after 2500 iterations run on 4 chains. For each stressor, we calculate total emissions and resource use as the sum of their off-farm (feed-associated) and on-farm components as described below.

# Off-farm (feed-associated) stressors

For off-farm stressors, we modelled the dry weight feed conversion ratio of study , , as:

such that and are the mean FCR for scientific name and and taxa group , respectively, and and are their respective standard deviations. To help with convergence we apply weakly-informative half-Cauchy priors on the scientific name and taxa group level ’s.

For off-farm stressors, we also modelled the proportions of feed originating from soy, crops, livestock, and fisheries. For each study i, Xi is a vector of feed proportions that sums to 1. We modelled this vector as:

such that is a shape parameter describing the distribution (uniform or degenerate) of the feed proportions for scientific name . We then reparameterized as:

such that is the vector of estimated feed proportions of scientific name and is the sample size (number of studies) for scientific name (Stan Development Team, 2020). To obtain our taxa group estimates, we modelled the estimated scientific name feed proportions as Dirichlet distributed, and reparameterized this as above to obtain a vector of estimated feed proportions for taxa group :

Finally, we calculated the species and taxa group level off-farm (i.e., feed-associated) stress () for each stressor (GHG, N, P, land, and water) by multiplying the feed requirements by the associated stressors of the feed, weighted by the feed composition:

Here, indexes the four feed ingredients (soy, other crops, fishmeal and fish oil, and livestock byproducts). is a constant that quantifies the stressors of each feed ingredient , while and are the posterior distributions of the estimated dry weight feed conversion ratio and the proportion of each feed component .

## On-farm stressors

On-farm nitrogen and phosphorus stressors are calculated from the same species and taxa-level means of and feed proportions, , described above for off-farm stressors. Here, on-farm N and P stress are estimated as the difference between the N and P content of each feed component and the species-specific N and P contents of each seafood product such that:

where and represent the nitrogen and phosphorus content of feed component and and represent the species-specific nitrogen and phosphorus content of a unit of fish or shellfish. While on-farm stress, , for N and P are both derived from the Bayesian means of FCR and the feed proportions, the other stressors, land, GHG, and water, are derived from nested means of the calculated stressor as described below. The non-feed associated land use refers to the land area allocated to the growth of a unit of output which applies to aquaculture systems that are ponds, recirculating systems and tanks. We calculated land stress (; m2a / tonne) as the reciprocal of annual yield (Y; tonnes / m2):

The on-farm greenhouse gas emissions () are calculated as the electricity use times the country-specific GHG emissions, plus the diesel, petrol, and natural gas use times each of their GHG stressor factors:

Here, indexes the four energy sources (electricity, diesel, petrol, and natural gas), represents the GHG emissions of energy source , and represents the energy use of source . While energy use does contribute to the other stressors, we only include it in the GHG stressor since the contribution to the other stressors is small. To calculate the on-farm water use, we estimated the evaporative losses over the surface area allocated to the unit of production as:

where represents the country-specific average surface evaporation rate and represents the grow-out period. We then model the Bayesian means for on-farm land, GHG, and water stressors with study nested in scientific name , nested in taxa group as shown below:

To help with convergence, we apply the following hyperpriors on all ’s at the scientific name and taxa group level. Specifically,

Finally, we apply priors on mean FCR at the taxa-level for all taxa groups except “miscellaneous diadromous fishes”:

We focus on FCR because it is a parameter found in all of our stressor models for which there is considerable information. Priors were derived from… Re-running the models with vs without these priors caused slight shifts in the posterior, but the major results (i.e., ordering of taxa results from highest to lowest stress) did not change.

# Missing Data Imputation

Missing data for FCR, feed proportions, yield, and all energy inputs (electricity, petrol, diesel, and natural gas) were imputed using taxa group as a categorical predictor and intensity and system type as ordinal predictors. In other words, the predicted value for missing data is pulled to the mean of other LCA studies of the same taxa group, intensity, and system type. For the feed proportions we used a Bayesian Dirichlet regression and for all other variables we used a Bayesian gamma regression with log link to estimate the effects of all predictors on each variable. For each missing data point, we used the median of their predicted posterior distribution as the imputed value for all subsequent analyses. All predictor variables are centered and scaled by two standard deviations. All models were fitted using Stan and implemented with brms in R. All models converged (i.e., diagnostics for all parameters showed 0.99 < < 1.01, > 0.1, and no divergent transitions) after 5000 iterations run on 4 chains. For the gamma regressions, the following hyperpriors were implemented on the coefficients (), intercept (), and shape () parameter for the gamma: