natureresearch

https://doi.org/10.1038/s41893-020-00668-1

Supplementary information

Identifying management actions that promote sustainable fisheries

In the format provided by the
authors and unedited

Supplementary Information for

Identifying Management Actions that Promote Sustainable Fisheries

Michael C. Melnychuk, Hiroyuki Kurota, Pamela M. Mace, Maite Pons, Cóilín Minto, Giacomo Chato Osio, Olaf P. Jensen, Carryn L. de Moor, Ana M. Parma, L. Richard Little, Daniel Hively, Charmane E. Ashbrook, Nicole Baker, Ricardo O. Amoroso, Trevor A. Branch, Christopher M. Anderson, Cody S. Szuwalski, Julia K. Baum, Tim McClanahan, Yimin Ye, Alessandro Ligas, Jilali Bensbai, Grant G. Thompson, John DeVore, Arni Magnusson, Bjarte Bogstad, Edward Wort, Jake Rice, Ray Hilborn

Correspondence to: mmel@u.washington.edu

This PDF file includes:

Supplementary Methods

Input data: stock status response variables....2
Input data: predictor variables....3
Data preparation....6
Correlation structures....8

Supplementary Discussion

Regional differences in management history....9

Supplementary Notes

Supplementary Note 1: Verification of model assumptions of temporal causality....11
Supplementary Note 2: Sensitivity analyses....13
Supplementary Note 3: Extended acknowledgments....15
Supplementary References....16
Supplementary Tables 3-7....17
Supplementary Figures 2-10....25

Other Supplementary Information for this manuscript includes:

suppl-table-1.pdf	(Supplementary Table 1. Management measures considered as predictor variables potentially affecting stock status)
suppl-table-2.pdf	(Supplementary Table 2. National/international-level fisheries management measures treated as interventions potentially affecting stock status)
suppl-table-8.pdf	(Supplementary Table 8. Summary of comparisons between sensitivity analyses and main run of results)
suppl-fig-1.pdf	(Supplementary Figure 1, with caption. Stock status history relative to the timing of fisheries management interventions for 288 individual stocks included in analyses—this is the full 289-page extension to the example shown in Fig. 2 of the main text.)

Supplementary Methods

Input data: stock status response variables

Stock assessment outputs are compiled for marine fish and invertebrate populations from around the world in the RAM Legacy Stock Assessment Database[21] (RAMLDB, version 4.491[22]). Assessments are usually conducted by government agencies, and key outputs are estimated time series of biomass (B, either total biomass or spawning stock biomass; commonly termed abundance) and fishing pressure (U, either an annual fraction harvested or an instantaneous fishing mortality rate; commonly termed exploitation rate). The start year considered in assessments is highly variable among stocks (Supplementary Table 4; Supplementary Figure 1). Some assessments further provide estimates of target reference points for biomass (B_{REF}) and fishing pressure (U_{REF}) that pair with biomass or fishing pressure time series. These target reference points are often based on maximum sustainable yield (MSY) or proxies for MSY[34] but may be based on other factors. For stocks with MSY-based reference points as well as actual target reference points provided in assessments, the actual targets were preferred. Pairing the time series with their respective target reference points, we obtain target ratios of $B/B_{\mathrm{REF}} = 1$ and $U/U_{\mathrm{REF}} = 1$. We note that the assumption of stationarity in reference points B_{REF} and U_{REF} is unlikely to be met for most stocks[27], however, the time series analysis involves annual changes in B/B_{REF} and U/U_{REF} rather than their magnitudes, and slow temporal variability in the reference points does not alter the direction of annual change in B or U.

For stocks that did not have target reference points provided in assessments, we fit surplus production models to catch and total biomass time series taken from assessments, similar to approach used previously[4,9,10,18,21]. Annual net surplus production values were calculated as the sum of annual catch and the change in total biomass from the current year to the following year, all in tonnes. We used a Pella-Tomlinson model[41] parameterised with B_{MSY} (with B as total biomass, TB), U_{MSY} (with U as harvest fraction, ER), and shape parameter γ to predict annual net surplus production ($\hat{S}$):

$$\hat{S} = \left(\left(\frac{\gamma}{\gamma - 1}\right) U_{MSY}\ \ B\right) - \left(\frac{U_{MSY}\ \ B^{\gamma}}{(\gamma - 1)\ \ {B_{MSY}}^{(\gamma-1)}}\right)$$

(Supplementary Equation 1)

For stocks that had a single missing reference point, the value for the other reference point was held fixed during the fitting procedure to estimate the single missing reference point. For stocks that had both reference points missing, both parameters were estimated simultaneously. Cross-validations with assessment-estimated reference points showed greater prediction accuracy for estimating both reference points when γ was fixed at the value 1.736, as previously estimated in a meta-analysis[42]. When only B_{MSY} was estimated (with U_{MSY} held fixed at the assessment-derived value), cross-validations showed greater prediction accuracy when γ was fixed at $\approx$1, which defines the Fox model[43]. When only U_{MSY} was estimated (with B_{MSY} held fixed at the assessment-derived value), cross-validations showed greater prediction accuracy when γ was specific to one of 13 taxonomic groups with values ranging from 0.65-2.43. These values were determined empirically by estimating a freely-varying γ while U_{MSY} and B_{MSY} were held fixed at assessment-derived values, and then calculating the arithmetic mean across the stocks in each taxonomic group.

A series of filters applied to surplus production model outputs guarded against poorly-estimated reference points. Similar filters were applied previously[4]. Estimated reference points were rejected if any of the following failures were observed:

(1) fewer than five years of annual net production and biomass were available
(2) estimated $U_{\mathrm{MSY}} < 0.005$
(3) estimated $U_{\mathrm{MSY}} > 0.85$
(4) estimated $B_{\mathrm{MSY}} < 0.07B_{\mathrm{MAX}}$, where B_{MAX} is the maximum recorded value in the total biomass time series
(5) estimated $B_{\mathrm{MSY}} > 2.085B_{\mathrm{MAX}}$
(6) the biomass range between 0 and the lesser of estimated carrying capacity, K, and B_{MAX} was divided into four equal intervals, and in the middle two intervals, three criteria were all required to fail in order for this filter (6) to be considered an overall failure: at least six net production values were negative; more than 50% of net production values were negative; and the sum of net production values was negative
(7) the calculated AICc value (Akaike's Information Criterion, adjusted for sample size) of the surplus production model fit was greater than any of three AICc values calculated for linear fits to annual surplus production values and biomass (fixed slope = 0 with freely-varying intercept; fixed intercept at origin with freely-varying slope; and both intercept and slope freely-varying)
(8) if B_{MSY} was available from the assessment and held fixed for estimating U_{MSY}, predicted surplus production at this fixed B_{MSY} was negative.

Numerical thresholds assumed for criteria 2-5 were based on ranges of values available from stock assessments. Typically, these filters 1-8 collectively exclude 13-19% of surplus production model fits (including fits for stocks that already had B_{REF} and U_{REF} available in assessments; these fits are still evaluated for use in cross-validations). Specifically for 83 stocks in our dataset without B_{REF} and/or U_{REF} from assessments, these filters collectively resulted in excluding B_{MSY} estimates for 13 of 55 stocks (24%) and excluding U_{MSY} estimates for nine of 63 stocks (14%); seven of these excluded stocks overlapped.

A sensitivity analysis ('Sensitivity 1—reference points') was conducted to examine the influence of including these *post-hoc* surplus production reference point estimates on observed results from later data analyses. Time series analyses (main Methods) were repeated after omitting these estimates, limiting the dataset to only stocks with reference points extracted from assessments. Comparisons of results with the main run are described in Supplementary Note 2.

Input data: predictor variables

There are 644 unique stocks contained in RAMLDB[22] with at least some available time series of B/B_{REF} or U/U_{REF}, including those with *post-hoc* estimated reference points. It was not feasible to collect management-related information for all of these, but we collected sufficient information for inclusion in analyses for 288 of these. In collecting management information, we ensured a high level of representation in terms of geography (Supplementary Table 3), taxonomic groups, population size, and fishing gears used. We focused mostly on stocks that are targeted in capture fisheries and that have been fished to such an extent that at some point in their history, fishing pressure had increased above U_{REF} or biomass had decreased below B_{REF}. These are typically the stocks of greatest management interest.

Fisheries management measures were considered at the stock level and at the national (or international) level as potentially influencing stock status. Stock-level measures consisted of two types of time series variables that were assembled by experts or during interviews with experts for each stock, and were occasionally supplemented with literature searches. First, the years in which a stock was under a formal rebuilding plan were assigned a '1', while all years not under a rebuilding plan were assigned '0'. Rebuilding plans vary in their duration after activation, are usually de-activated following stock recovery, and may later be re-activated as deemed necessary for rebuilding (Supplementary Figure 1). Rebuilding plans are commonly implemented when a stock's relative biomass B/B_{REF} is estimated to be 'too low' or below some threshold such as 0.5 (Supplementary Table 4). Rebuilding plans were the only management measure considered to have potential influence only in the years in which they were active (Supplementary Table 1). Second, similar to a previous approach[29], an aggregate variable of stock-level management intensity ranged from 0-1 and comprised five components other than rebuilding plans: scientific surveys of fish abundance; stock assessments; harvest control rules; fleet-wide catch limits; and individual quotas. The year in which each of these measures was first implemented for a stock incremented the aggregate index by 1/5. Any order of the five components was allowed, and if two components were implemented in the same year, the index incremented by 2/5 in that year. Unlike rebuilding plans, these other measures were treated as having a potential influence that persists indefinitely after they were first implemented (Supplementary Table 1). All years during and after the first use of these components were considered to be potentially influential, so (in contrast to how rebuilding plans were treated) the aggregate index increases monotonically. For example, if scientific surveys were implemented in some year and then ceased in some later year, or if they are only conducted every few years, the aggregate index does not decrease after first usage. Rebuilding plans were not considered as part of the aggregate index because their effect on stock status is expected to occur only in years in which they are active.

Similar to the stock-level aggregate index of management intensity, an aggregate variable of national-level management intensity ranged from 0-1, increased monotonically, and comprised three components, each of which incremented the aggregate index by 1/3: country-specific declaration of an Exclusive Economic Zone (EEZ)[15]; country-specific ratification of either the United Nations Food and Agriculture Organisation Compliance Agreement (UNCA)[16] or the United Nations Fish Stocks Agreement (UNFSA)[17], whichever was ratified first; and implementation of a major fisheries policy considered to have potential influence on most or all stocks in the country or region. Examples of this major fisheries policy included the U.S. Sustainable Fisheries Act and the European Union's 2002 reform of the Common Fisheries Policy. For stocks managed under tuna Regional Fisheries Management Organisations (tRFMO), this major fisheries policy consisted of the convention that governs the tRFMO. For stocks that are fished by multiple countries (e.g., West African stocks), management measures were specific to the country with the greatest proportion of catch of the stock. For tuna stocks, the year of first ratification of a UN agreement was likewise based on the country with the greatest proportion of catch of the stock. The national/international-level fisheries policies considered in analyses are listed in Supplementary Table 2 along with their year of implementation.

Rebuilding plans, stock-level management measures, and national/international-level management measures tended to co-vary in their usage. There were few stock:years that had low stock-level management intensity and high national-level management intensity together, and few stock:years that had the opposite (Supplementary Figure 4a). Rebuilding plans did not occur when stock-level management intensity was 0, and rarely occurred when national-level management intensity was 0. Rebuilding plans were most commonly activated when both management indices were high, although there were also cases in which one or the other management index was low or intermediate while rebuilding plans were in place (Supplementary Figure 4a). Across all stocks, EEZs were in place for 72% of stock:years, individual quotas were in place for 23% of stock:years, and other management measures were intermediate between these proportions (Supplementary Figure 4b). Considering only the years while under active rebuilding plans, the proportions ranged from 94% of stock:years (for EEZs) to 43% of stock:years (for individual quotas; Supplementary Figure 4b), suggesting that rebuilding plans tended to be used after various other measures had already been implemented. Years of implementation of stock-level measures and national-level measures are shown in Supplementary Figure 1 for each stock, and summarised in Supplementary Table 4. Regional changes over time in the implementation of individual management measures and aggregate indices of management intensity are shown in Supplementary Figure 2 for each region and in Supplementary Figure 3 for each measure.

Management measures were treated either as a Boolean variable (rebuilding plan) or as incremental indices in analyses, but in reality, they represent a continuum. For example, some harvest control rules would be expected to have greater effect on stock status than other harvest control rules[19], particularly when output harvest recommendations are backed by law instead of being discretionary[9]. Some rebuilding plans are stronger than others, ranging from complete fishery closures to temporary, modest reductions in fishing pressure[20,24]. Some stock assessments provide more accurate estimates of stock status than others[24], which affect the basis on which management decisions are made. While such nuances are frequent and may realistically influence stock status differently, it was necessary to make simplifying assumptions when categorising management measures consistently across diverse regions and stocks.

Life-history traits and taxonomic groups were also considered as potentially influencing stock status. Life-history traits initially considered included: (1) natural mortality rate, M; (2) age at 50% maturity, A_{M50}; (3) length at 50% maturity, L_{M50}; (4) maximum age, A_{MAX}; (5) maximum length, L_{MAX}; (6) von Bertalanffy growth, κ; and (7) trophic level, TL. Variables 1-3 were available for some individual stocks; otherwise, values from a nearby stock were assumed, and if still not available, average values at the global species level were extracted from FishBase[44] or SeaLifeBase[45]. Paired scatterplots showed strong correlation between several pairs of variables in either linear (Supplementary Figure 7) or log space. To avoid problems with collinearity, variance inflation factors were calculated[38] and only two life-history variables were carried forward into regression analyses, A_{M50} and L_{MAX}. A categorical variable representing broad taxonomic groups (demersal fish; pelagic fish; invertebrates) was also considered as a predictor variable.

Two fishery-related attributes were considered as predictor variables potentially influencing stock status. First, a categorical variable distinguishing single-species fisheries from mixed-species fisheries was considered for each stock. In cases where some fleets catch the stock alone while other fleets catch the stock in a mix of species, the variable was assigned according to the principal fleet. Second, the product of MSY and average ex-vessel price for a given stock represented its Maximum Sustainable Landed Value ($MSLV$), as quantity and price together drive incentives for targeting by fishing fleets[23]. If an estimate of MSY was not available for a stock (which was the case for only 6.6% of stocks), the mean catch across the full time series with leading zeros removed was instead used (Supplementary Figure 1). These values of MSY and mean catch were highly correlated for the stocks that had both values available ($r = 0.964$). Predicted prices were generated from an external mixed-effects regression model fit to observed ex-vessel prices from national price datasets. This provided predicted prices even for stocks without observed prices, based on their nested taxonomic levels and regional covariates. The mean price during 2001-2010 was calculated and was multiplied by MSY (or mean catch) to obtain *MSLV*. Time series of catches, observed and predicted prices, as well as estimates of MSY, mean catch, and mean predicted price for years 2001-2010 are shown in Supplementary Figure 1.

For plotting, stocks were assigned to regions based on their geographic distributions and management authorities. The 288 sampled stocks for data analyses were distributed among 17 regions, with 3-31 stocks per region (Supplementary Table 3; Supplementary Figures 1-2).

Data preparation

Management measures and other predictor variable data were collected for 296 stocks, but U/U_{REF} or B/B_{REF} response variable data were available for only 288 of these. Reference point estimates were drawn from stock assessments for 232 stocks (U_{REF}) and 240 stocks (B_{REF}), and were drawn from surplus production model fits for 54 stocks (U_{REF}) and 42 stocks (B_{REF}). The other eight of the originally-available 296 stocks were excluded because they did not have available reference points after applying the set of filters described above to surplus production model estimates. The 288 stocks included in analyses had between 1-67 years of U/U_{REF} and/or B/B_{REF} estimates available (mean 41.8 years for U/U_{REF} and 43.1 years for B/B_{REF}). Stocks with <10 years of available data were excluded from analyses. This yielded a total of 11,944 stock:years of U/U_{REF} estimates and 12,162 stock:years of B/B_{REF} estimates across all stocks. Missing values within otherwise contiguous time series existed for nine stock:years of B/B_{REF} (across four stocks) and one stock:year of U/U_{REF}; these few missing values were linearly interpolated.

In the regression models described in the main Methods, the potential effect on stock status of most management measures was assumed to be persistent following the implementation of a measure. Two management-related variables (stock-level aggregate index of management intensity; and national/international-level aggregate index of management intensity) were considered to potentially influence stock status during their year of implementation and all years following in the stock's time series. In contrast, the third management-related variable, 'under rebuilding plan', was considered to potentially affect stock status only during the specific year(s) in which it was active (Supplementary Figure 1). The rebuilding plan effect was separated into two components, an immediate component (in the first year of implementing a rebuilding plan) and a persistent component (for all years after the first year, until the rebuilding plan was de-activated or until the end of the time series). This separation of components, described further in the main Methods, allowed for distinguishing immediate effects from longer-term effects. In particular, fishing pressure is likely to decrease immediately after activating a rebuilding plan (because fishing fleets can respond to management changes within the same year or fishing season), whereas biomass may require several years under a rebuilding plan before starting to recover (because of biological constraints on rates of population increase).

Stock time series were partitioned into (up to) two distinct phases based on values of U/U_{REF}, B/B_{REF}, and catch/MSY or catch/(mean catch). This allowed us to focus our research questions (about management influences on fishing pressure and biomass) on the most applicable portion(s) of a stock's available time series. The first phase, 'developing fishery', was considered from the start of a stock's available time series until any of the following criteria were met: (a) $B/B_{\mathrm{REF}} < 0.8$; (b) $U/U_{\mathrm{REF}} > 1$; (c) catch/MSY > 1; (d) catch/(mean catch) > 1.25; or (e) a rebuilding plan was implemented. To guard against misclassifications arising from truncated time series (if data were available only after actual fishery development), if either of the two following conditions were observed, the stock was assumed to already be past the ‘developing’ phase: (f) $B/B_{\mathrm{REF}} < 1$ in the first year of the available time series; or (g) catch in the first year when both B/B_{REF} and U/U_{REF} data were available was less than 1.25 times the mean catch in the previous ten years (which may occur if catch time series extend further into the past than B or U estimates provided in assessments). The second phase, ‘mature fishery’, was assumed to include all years after the ‘developing’ phase. For stocks lacking a ‘developing’ phase in their available time series, the full time series was considered to be ‘mature’.

Stocks in some regions tended to pass from ‘developing’ to ‘mature’ fishery phase sooner than stocks from other regions. For example, among the regions considered, fisheries in Australia, New Zealand, West Africa, South America, and tRFMOs tended to develop later than fisheries in other regions (Supplementary Figure 1). In contrast, declaration of EEZs and ratification of UN agreements tended to cohere more closely in time across regions (Supplementary Figures 2 and 3). This implies that EEZ declaration and ratification of UN agreements would generally be well into the ‘mature’ fishery phase for stocks in earlier-developing regions, but may be either in the ‘developing’ phase or early in the ‘mature’ phase for stocks in later-developing regions. If the implementation of a management measure occurred prior to the beginning of a stock’s ‘mature’ fishery phase, it would have no influence on analyses that were restricted to the ‘mature’ fishery phase.

We recognise that other factors besides management actions, such as environmental conditions, may influence stock abundance and therefore affect the timing of transitions from the ‘developing’ phase into the ‘mature’ phase. Despite these possible external influences, the blocking of time series data into ‘developing’ and ‘mature’ phases allows for some analyses to be focused solely on the ‘mature’ phase, when implementation of management measures is most relevant (main Methods sections ‘*Base model for stock status trends*’ and ‘*Predicting short-term responses to management*’). In other analyses, when a greater range of magnitudes of U/U_{REF} and B/B_{REF} was necessary (main Methods section ‘*Predicting equilibrium responses to management*’), the full time series including the ‘developing’ phase was considered. Classified phases are shown for all stocks in Supplementary Figure 1. The start years of phases, and status of U/U_{REF} and B/B_{REF} at the time of these phase starts, are summarised in Supplementary Table 4. Some ‘mature’ phases were only a few years in duration, so to guard against small sample sizes, all analyses described below required a minimum of 10 years of data in the ‘mature’ phase for a given stock and response variable. This filtered out 1 stock for U/U_{REF} only, 1 stock for B/B_{REF} only, and 1 stock for both U/U_{REF} and B/B_{REF}.

A sensitivity analysis (‘Sensitivity 2—time series length’) was conducted to examine the influence on observed results of this 10-year threshold for inclusion. Data analyses were repeated, instead requiring a minimum of 20 years of data for a given stock and response variable. Comparisons of results with the main run are described in Supplementary Note 2.

Response variables U/U_{REF} and B/B_{REF} were ln-transformed to ensure symmetrical proportional changes above and below target ratios of 1 (e.g., a doubling from $U/U_{\mathrm{REF}} = 1$ to 2 is symmetrical with a halving from $U/U_{\mathrm{REF}} = 1$ to 0.5). Response variables were subsequently differenced for time series regression analysis to achieve stationarity[36]. First-order differences were determined to be sufficient for most stocks (see next section). Numerical predictor variables for regression analyses (A_{M50}, L_{MAX}, and $MSLV$) were centred by subtracting the arithmetic mean and standardised by dividing the result by the standard deviation. Model fit diagnostics were evaluated, and are reported in Supplementary Table 7 for both response variables, $\Delta\ln(U/U_{\mathrm{REF}})$ and $\Delta\ln(B/B_{\mathrm{REF}})$.

Correlation structures

ARIMA (autoregressive integrated moving average) model correlation structures contain components for autoregression (p), differencing (d), and moving average prediction errors (q) for a univariate time series. The appropriate orders of p, d, and q can be determined for a given time series through statistical tests for stationarity, inspection of autocorrelation function plots and partial autocorrelation function plots, or evaluating criteria for statistical fitting[36]. The auto.arima() function of the R package 'forecast'[46] combines several of these checks to provide an optimal set of parameters for a given time series. In a hierarchical model with multiple time series, however, the same orders of p, d, and q must be assumed across all groups (using the same grouping structure as for random effects[37]) even though the optimal set of parameters may vary among individual groups. To identify the best overall set of p, d, and q parameters across stocks, we used the auto.arima() function[46] to identify the best set of parameters for each stock in the 'mature' fishery phase (as well as for the full time series, including the 'developing' phase), and then we summarised these best-identified sets across stocks to reveal an overall best set.

First, we evaluated the order of required differencing to ensure stationarity. For $\ln(U/U_{\mathrm{REF}})$, the most frequent order of required differencing was 1 for both the 'mature' phase and for the full time series (Supplementary Table 5). The next-most-frequent required order was 0; few stocks required second-order differencing. For $\ln(B/B_{\mathrm{REF}})$, the most frequent required order of differencing was also 1, followed by order 2 for both the 'mature' phase and full time series (Supplementary Table 5). This most frequent order of 1 was assumed for all analyses ($d = 1$).

Second, using first-order differenced time series, the order of autoregressive and moving average components required to minimise AICc were identified[46]. For $\ln(U/U_{\mathrm{REF}})$, the most frequent combination of parameters was an ARIMA(0,1,0) structure in the 'mature' phase as well as for the full time series (Supplementary Table 6). However, there were also several stocks for which the best-fit structure required 1 or 2 orders of p (with $q = 0$), or 1 or 2 orders of q (with $p = 0$). Few stocks required >0 orders of p and >0 orders of q simultaneously. For $\ln(B/B_{\mathrm{REF}})$, the most frequent combination of parameters was an ARIMA(1,1,0) structure, i.e., lag-1 autoregression, in the 'mature' phase as well as for the full time series. However, orders of $p = 0$ or 2 were also relatively frequent, as were orders of $q = 1$ (Supplementary Table 6). Because most stocks had best-fit orders of $p = 0$ or 1 and $q = 0$ or 1, an ARIMA(1,1,1) structure was selected for the main analysis, erring on the side of including additional parameters that may be unnecessary for some stocks (rather than failing to include additional parameters that may be necessary for other stocks). The selected ARIMA(1,1,1) structure was applied, with calendar year treated as the time covariate and stock as the grouping variable[37]. To correspond with this grouping structure, stock was also treated as a random intercept in regression models[37]. In this ARIMA structure with $p = 1$ and $q = 1$, there is one autoregressive parameter ϕ and one moving average parameter θ to estimate, respectively. The magnitudes of the estimated values of these parameters are indicative of whether the fitted model is considered to be temporally causal (see Supplementary Note 1).

Four sensitivity analyses ('Sensitivity 3a,b,c,d—ARIMA structure) were conducted to examine the influence of ARIMA model structure assumptions (i.e., the selected values of p and q) on observed results. Analyses were repeated assuming alternative structures of ARIMA(0,1,0), ARIMA(1,1,0), ARIMA (2,1,0), and ARIMA(0,1,1). Comparisons of these sensitivity results with those of the main run are described in Supplementary Note 2.

Supplementary Discussion

Regional differences in management history

Management measures at the stock level and at the national (or international) level considered in analyses are defined in Supplementary Table 1. Summarising all stocks in our analysis (which comprise assessed stocks primarily from regions with high research and management capacity), Fig. 1 shows the history of implementing these management measures since 1950. While these management measures have all been increasingly used over this period, considerable variability exists among regions and individual stocks in whether and when these measures have been applied. Variability among regions also exists in the timing of implementing management measures with respect to the timing of when stocks transitioned from their 'developing' fishery phase to 'mature' fishery phase (Supplementary Figure 1). For example, the declaration of EEZs and ratification of UN agreements tended to occur prior to the transition into the 'mature' fishery phase for stocks in some regions, but well after the transition for stocks in other regions. Because the timing of changes in mean stock status differed among regions[4], while the timing of implementing EEZs and UN agreements tended to be synchronous across regions (Supplementary Figures 2-3), this implies that regions differ in their patterns of management history with respect to stock status history. The history of implementing management measures for individual stocks is shown in the multi-page Supplementary Figure 1. Grouping stocks by region, Supplementary Figures 2 and 3 show how temporal patterns of implementing management measures have varied among regions.

At the stock level, rebuilding plans or catch moratoria have been in place in any given year for nearly half the studied stocks in the United States[20,34], Canada, New Zealand[47], and South Africa (Supplementary Figures 2 and 3). In contrast, they have not been applied for any stocks in our dataset from the Mediterranean or Black Seas ('Europe–Med/Black Sea') or West Africa despite a long history of overfishing in these regions[4,5] (Supplementary Figure 3). Regions with at least some history of implementing rebuilding plans include European Union waters of the northeast Atlantic[25] ('Europe(EU) NE Atl', including Atlantic Ocean, North Sea, and Baltic Sea, but not the Mediterranean), Australia[48], Japan[49], South America, tuna RFMOs[28], Russia East Coast, and European (but non-European Union) waters of the northeast Atlantic ('Europe(non-EU) NE Atl'), consisting of Norway[50], Iceland, the Faroe Islands, and some transboundary stocks shared with Russia (Supplementary Figures 2 and 3).

Among the other five stock-level management measures, scientific surveys tended to be implemented earliest in most regions, with the exception of Europe(non-EU) NE Atl, Australia, and tuna RFMOs (where formal stock assessments and fleet-wide catch limits were generally implemented earlier), Russia East Coast (where fleet-wide catch limits were implemented earlier), and New Zealand and South Africa (where fleet-wide catch limits and individually-allocated catch quotas were implemented earlier; Supplementary Figures 2 and 3). In US regions, individual quotas tended to be implemented most recently. In other regions (Canada East Coast, Europe(non-EU) NE Atl, Europe(EU) NE Atl, Russia East Coast, Australia, and New Zealand), harvest control rules tended to be implemented most recently. Regions with limited use of harvest control rules tended to have the highest relative fishing pressures in recent years (Supplementary Figure 3). While most regions have by now implemented these five stock-level management measures for at least half the assessed stocks, some management measures have been implemented for no or few stocks in other regions. These include Europe–Med/Black Sea (no individual quotas; few fleet-wide catch limits or harvest control rules, although as of 2019 maximum allowable fishing efforts have been established for the trawl fleets targeting 29 demersal stocks in the West Mediterranean and Adriatic Seas), West Africa (no harvest control rules), South America (no harvest control rules), Japan (no individual quotas), US Northeast and Southeast (few individual quotas), Canada East Coast (few harvest control rules), and tuna RFMOs (few harvest control rules, individual quotas, or scientific surveys; Supplementary Figures 2 and 3).

At the national (or international) level, management measures are typically applied simultaneously across most stocks in a country or region, so the changes over time are more discrete compared to the more gradual implementation of stock-level management measures (Supplementary Figures 2 and 3). There also tends to be less variability among regions in the implementation of national management measures, as many involve international agreements that in most cases were ratified by countries around the same time[16,17]. Most countries declared EEZs[15] in the late 1970s, with the exception of South America (Chile and Peru declared earlier, in 1947; Supplementary Table 2), Europe–Med/Black Sea (France and Spain have declared EEZs in the Mediterranean, but only recently), and tuna RFMOs (these stocks are highly migratory and are typically distributed across an ocean basin, so EEZs are less relevant). The ratification of the UN Compliance Agreement[16] or the UN Fish Stocks Agreement[17], whichever was first ratified by a country, occurred around the same time across regions, between the mid-1990s to early 2000s (Supplementary Figures 2 and 3; Supplementary Table 4). These international agreements are specifically related to the high seas or illegal fishing for cross-boundary stocks, but they may also exert an indirect influence on stocks in national or sub-national waters as stronger fisheries management commitments at the international level may permeate down into national and sub-national management systems. The implementation of other major pieces of fisheries legislation at the national or international level was more variable across regions (Supplementary Table 2). Conventions for some tuna RFMOs were established in the 1950s or 1960s (IATTC, ICCAT) while others were not established until the 1990s or 2000s (CCSBT, IOTC, WCPFC). At the national level, most of the key pieces of fisheries legislation thought to potentially affect stocks were implemented in the 1980s (New Zealand[47], Canada, Norway[50]) or 1990s (Iceland[51], Australia[48], Chile, Peru, Faroe Islands[52], US[34], South Africa, Argentina), while others were implemented more recently in the 2000s (EU[2,25,53] and Russia). Comparable pieces of major fisheries legislation (Supplementary Table 2) have not been implemented in Japan, Europe–Med/Black Sea (before 2016, but after which new demersal management plans have been adopted), or West Africa.

The five stock-level management measures and three national/international-level measures described above were modelled as remaining in place after their initial implementation. This is a reasonable assumption for the vast majority of measures that have been applied to stocks. One rare exception occurred for Faroe Plateau Atlantic cod, for which a quota system was implemented in 1994, but remained in place only until 1996, when a system of individual effort allocations was implemented in its place[52]. Other temporary interruptions in the use of a management measure are more common, for example if a stock assessment or a scientific survey is not carried out every year, but only periodically.

Aggregate indices of stock-level management intensity (comprising use of scientific surveys, stock assessments, harvest control rules, fleet-wide catch limits, and individual quotas) and of national-level management intensity (comprising EEZ declaration, first ratification of UN Compliance Agreement or UN Fish Stocks Agreement, and a major piece of national fisheries legislation) are also variable among regions in their timing and prevalence (Supplementary Figures 2 and 3). Stock-level management intensity increased gradually over several decades while increases in national-level intensity were generally more punctuated as EEZs were declared and international agreements were ratified. Sums in Supplementary Figure 2 and proportions in Supplementary Figure 3 reflect the number or proportions of stocks as well as the values ranging from 0-1 for each individual stock. For example, in these aggregate sums or joint proportions, a value of half the number of stocks could result from half of the stocks at value 1, from all of the stocks at value 0.5, or any such combination. These aggregate sums or joint proportions are currently at least three quarters of the number of stocks in most regions, with the exception of Europe–Med/Black Sea (both indices at about half the number of stocks), Japan and West Africa (both indices at about two thirds the number of stocks), and tuna RFMOs (stock-level index at about one third the number of stocks and national-level index at about two thirds the number of stocks). Two of these four regions have the highest median U/U_{REF} among all regions, Europe–Med/Black Sea (2.3) and West Africa (1.6) (Supplementary Figure 3).

Most regions shown in Supplementary Figures 2 and 3 contain a high proportion of stocks that would typically be considered as data-rich. At least some of the stocks from Europe–Med/Black Sea[5], West Africa, and tuna RFMOs[6,11] may arguably be considered to have intermediate levels of data availability, with limited use of scientific surveys or application of relatively simple stock assessment methods. Due to the lack of available scientific estimates of fishing pressure or abundance relative to reference points, this study does not include stocks that would typically be considered as data-poor, which tend to occur disproportionately in developing countries. Owing to limited financial resources, these same regions tend to have more limited capacity in their fisheries management systems[7,13]. These differences in overall capacity among regions are likely to affect how effective any given management measure may be. Our analysis evaluated overall effects of management measures on trends in stock status across all regions simultaneously, even though those regions differ in management capacity. Future work could involve more detailed analyses of how the effectiveness of any given management measure (i.e., its influence on stock status trends) may vary among regions that differ in financial and management capacity.

Supplementary Notes

Supplementary Note 1: Verification of temporal causality

The assumption of temporal causality in ARIMA models can be verified from estimated parameters for autoregressive ($\hat{\phi}_1, \ldots, \hat{\phi}_p$) and moving average ($\hat{\theta}_1, \ldots, \hat{\theta}_q$) processes. We applied an ARMA(1,1) model to first-order differenced time series, which is equivalent to applying an ARIMA(1,1,1) model to un-differenced time series. In both cases, there is only one ϕ parameter and one θ parameter to estimate.

Temporal causality applies to the autoregressive component of ARMA or ARIMA models. The model can be said to be temporally causal if the time series can be written as a one-sided linear process, i.e., if the present state of the response variable depends on past-year states but not on future-year states[54]. For an ARMA(1,1) model, if the following conditions are met, then temporal causality can be verified:

- $\hat{\phi} < 1$
- there are no common roots between autoregressive and moving average polynomials
- response variable time series are stationary

These criteria were all met. For the base model (Equation 1) and coupled-variable model (Equation 3), the following parameter estimates were observed:

Equation 1, response variable $\ln(U/U_{\mathrm{REF}})_{t\rightarrow t+1}$: $\hat{\phi} = 0.55$; $\hat{\theta} = -0.80$
Equation 1, response variable $\ln(B/B_{\mathrm{REF}})_{t\rightarrow t+1}$: $\hat{\phi} = -0.46$; $\hat{\theta} = 0.60$
Equation 3, response variable $\ln(U/U_{\mathrm{REF}})_{t\rightarrow t+1}$: $\hat{\phi} = 0.50$; $\hat{\theta} = -0.71$
Equation 3, response variable $\ln(B/B_{\mathrm{REF}})_{t\rightarrow t+1}$: $\hat{\phi} = 0.16$; $\hat{\theta} = -0.03$

The parameter estimates were different within each model, so there were no common factors, and therefore no common roots. Overall across stocks, first-order differencing was sufficient to ensure stationarity (Supplementary Table 5). Therefore, the assumption of temporal causality was verified. Analyses weighted by *MSLV*, and analyses involving disaggregated management measures instead of aggregate indices of management intensity, also met the above criteria.

These parameter estimates of the ARIMA correlation structure reveal information about the time series of stocks included in our analysis. After differencing time series to ensure stationarity, autoregressive parameter estimates were between 0-1 for $\Delta\ln(U/U_{\mathrm{REF}})$ for both the base model and coupled-variable model, indicating that unexplained changes tend to persist, but dampen, over time. For $\Delta\ln(B/B_{\mathrm{REF}})$ in the base model, the autoregressive parameter estimate was $-1 < \hat{\phi} < 0$, indicating that unexplained changes also tended to dampen but flip from positive to negative and back. This may result from irregular recruitment anomalies affecting biomass changes over time. For $\Delta\ln(B/B_{\mathrm{REF}})$ in the coupled-variable model, $\hat{\phi}$ was weakly positive, thus incorporating the strong influence of U/U_{REF} magnitude on $\Delta\ln(B/B_{\mathrm{REF}})$ reduced the relative importance of autoregression on changes in biomass. Moving average parameter estimates generally followed opposite patterns as those for autoregressive terms. The negative moving average estimates for $\Delta\ln(U/U_{\mathrm{REF}})$ may indicate that, whether (Equation 3) or not (Equation 1) B/B_{REF} magnitude is explicitly accounted for in the regression model, changes in U/U_{REF} tend to over-compensate for ‘prediction errors’ in previous years. For example, higher-than-expected levels of U/U_{REF} in some year may require adjustment downwards in the following year. For $\Delta\ln(B/B_{\mathrm{REF}})$ in the base model, prediction errors tended to propagate but dampen over time, whereas in the coupled-variable model with U/U_{REF} magnitude explicitly included, the moving average parameter estimate greatly weakened.

We note that this verification of temporal causality refers to the correlation structure of the response variables. This verification does not imply causal effects of predictor variables on the response variables. As with regression models in general, in ARIMA models, predictors are assumed to be independent of the responses. In reality, however, experimental or random implementation of management measures is rare. Management measures are often implemented in response to changing stock status. Although we were not able to control for the non-random implementation of management measures, ARIMA models do at least distinguish changes in stock status that occur in the years prior to implementing a management measure from changes that occur in the years during and after implementation. In other words, these models allow for separating the baseline trends in U/U_{REF} and B/B_{REF} (pre-implementation) from the impacted trends (post-implementation) as a stock switches between treatment groups.

Supplementary Note 2: Sensitivity analyses

Four types of sensitivity analyses (eight runs in total) were conducted to evaluate how assumptions or model structures used in the main run potentially affected observed results. For evaluating results, we use two approaches for comparing outputs of sensitivity runs to those of the main run. First, we focus on coefficient estimates of the four management-related parameters (including both components of rebuilding plans) estimated under the base model, which are shown in Fig. 3 for the main run. Second, we visually compare Figs. 3-5 and Supplementary Figure 8 (Fig. 4 is a subset of Supplementary Figure 8) to their counterparts produced under sensitivity runs to evaluate whether any of the changes to assumptions or model structures result in different conclusions drawn from observed results. The four types of sensitivity analyses considered were:

1. in the main run, if reference points B_{REF} or U_{REF} were not provided in assessments, we estimated these *post-hoc* with a surplus production model (Supplementary Equation 1). In this 'Sensitivity 1—reference points', we do not estimate missing reference points, we use only the reference points provided in assessments, which reduces the sample size for analyses.
2. in the main run, we required a minimum of 10 years of B/B_{REF} or U/U_{REF} data for inclusion in time series analyses. In this 'Sensitivity 2—time series length', we instead require a minimum of 20 years of data, which also reduces the sample size.
3. in the main run, we used an ARIMA(1,1,1) correlation structure for ARIMA(p, d, q), which contains one autoregressive parameter, p, and one moving average prediction error parameter, q. In this 'Sensitivity 3—ARIMA structure', we consider four alternative correlation structures that were commonly observed to be the best parameter set for some stocks (Supplementary Table 6). These four alternative correlation structures differ in the number of parameters included for p and q:
   (3a) ARIMA(0,1,0)
   (3b) ARIMA(1,1,0)
   (3c) ARIMA(2,1,0)
   (3d) ARIMA(0,1,1).
4. in the main run, we weighted individual stocks equally. We also considered an alternative weighting scheme, in which stocks were weighted by their mean MSLV; results under this alternative stock-level weighting scheme are shown in Supplementary Figures 6 and 8. In this 'Sensitivity 4—regional weighting', we consider two regional-level weighting schemes. Regression weights are still applied to individual stocks, but in the following regional-level weighting schemes, all the stocks from a given region share the same weight:
   (4a) regional weights proportional to the number of stocks in RAMLDB with available time series of U/U_{REF} or B/B_{REF}. Regions were assigned weights of $n_{\mathrm{r,RAMLDBfull}}/n_{r,\mathrm{paper}}$, where $n_{r,\mathrm{paper}}$ is the number of stocks in region r included in this analysis, and $n_{\mathrm{r,RAMLDBfull}}$ is the number of stocks contained in RAMLDB with any available estimates of U/U_{REF} or B/B_{REF}. Stock counts for $n_{r,\mathrm{paper}}$ and $n_{\mathrm{r,RAMLDBfull}}$ are listed in Supplementary Table 3.
   (4b) regional weights proportional to the number of stocks in RAMLDB that at some point in their time series had $U/U_{\mathrm{REF}} > 1$ or $B/B_{\mathrm{REF}} < 1$. Regions were assigned weights of $n_{\mathrm{r,RAMLDBsub}}/n_{r,\mathrm{paper}}$, where $n_{\mathrm{r,RAMLDBsub}}$ is the subset of stocks in region r contained in RAMLDB that at some point in their history met at least one of these conditions for U/U_{REF} or B/B_{REF}.

These two sets of regional-level weights were each scaled by their median, such that the median weight among the 17 regions was 1. Values of sample weights are listed in the file "regional-weights.csv", one of the input files provided with code for reproducing analyses.

In general, none of the sensitivity analyses led to different take-away conclusions than those from the main results. Certain differences were observed, which are summarised in Supplementary Table 8 and described further below:

1. For 'Sensitivity 1—reference points', three of four estimated coefficients of management-related parameters for $\Delta\ln(U/U_{\mathrm{REF}})$ and all four estimated coefficients for $\Delta\ln(B/B_{\mathrm{REF}})$ were similar to those from the main run (Supplementary Table 8). The one notable difference was that the coefficient for $b_6 Reb_{immediate_{t,j}}$ in 'Sensitivity 1—reference points' was only about one quarter the magnitude of its counterpart in the main run (Supplementary Table 8). Despite this weaker effect compared to the main run, it was still the strongest effect overall in 'Sensitivity 1—reference points' (as seen in the Fig. 3 counterpart), so still resulted in decreased fishing pressure in the first year of implementing a rebuilding plan (as seen in the Fig. 4 and Supplementary Figure 8 counterparts). This immediate decrease only reduced U/U_{REF} partway to target levels of 1 (not all the way, as seen in the main run); after the first year, the decrease in U/U_{REF} continued at a slower rate, reaching target levels before the end of the 10 year in the medium and high management intensity scenarios of the counterpart to Supplementary Figure 8. Some differences were observed in equilibrium predictions (Fig. 5) between 'Sensitivity 1—reference points' and the main run. At high levels of management intensity, observed results were similar to those from the main run, although mean B/B_{REF} never exceeded 1 and mean U/U_{REF} never decreased below 1 even at management intensity values of 1. The threshold at which rebuilding plans activated occurred at slightly lower levels of management intensity in 'Sensitivity 1—reference points', and the proportion of years spent under rebuilding plans was generally greater than in the main run.
   Because reference point estimates for U_{REF} and B_{REF} from surplus production models generally show limited bias (based on cross-validations with estimates drawn from stock assessments), the differences outlined above between 'Sensitivity 1—reference points' and the main run are likely due to the subset of stocks excluded (sample size in the sensitivity analysis was reduced by 20-27% compared to the main run). Median U/U_{REF} across all stocks and years was nearly identical for the main run (0.975) and 'Sensitivity 1—reference points' (0.977), and median B/B_{REF} was similar (main run, 1.024; 'Sensitivity 1—reference points', 1.095).
2. For 'Sensitivity 2—time series length', estimated coefficients of management-related parameters were all similar to those from the main run (Supplementary Table 8). Visual comparisons with Figs. 3-5 and Supplementary Figure 8 revealed no notable differences between the main run and 'Sensitivity 2—time series length'.
3. For 'Sensitivity 3—ARIMA structure', estimated coefficients of management-related parameters were all similar to those from the main run (Supplementary Table 8). Visual comparisons with Figs. 3-5 and Supplementary Figure 8 revealed no notable differences between the main run and 'Sensitivity 3a,b,c,d—ARIMA structure'.
4. For 'Sensitivity 4—regional weighting', both weighting schemes (a, b) which involved weighting regions in proportion to numbers of assessed stocks by region had estimated coefficients of management-related parameters similar to those from the main run (Supplementary Table 8), and visual comparisons with Figs. 3-5 and Supplementary Figure 8

revealed no notable differences between the main run and ‘Sensitivity 4a,b—regional weighting’.

Supplementary Note 3: Extended acknowledgments

The following individuals graciously and patiently provided management-related information about fisheries included in the analysis. This work would not have been possible without their expertise.

US West Coast/Alaska: Andi Stephens, Owen Hamel, Jennifer Ford, Mary Furuness, Kevin Duffy, Joshua Lindsay, Sean Matson, Steve Ralston, James Thorson, Chantel Wetzel
US East Coast: Myra Brouwer, Ryan Rindone, Steve Atran, Karyl Brewster-Geisz, Stephen Holliman, Rick Hart, Mike Travis, Jose Montanez, Peter Hood, Kari MacLauchlin, Jeannette Banobi
Canada West Coast: Jaclyn Cleary, Kristen Daniel, Karen Dwyer, Brian Healey, Roger Kanno, Robert Tadey, Dan Clark, Ken Fong, Wellsley Hamilton, Kim Hardacre, Adam Keizer, Allen Kronlund, Anna Magera, Guy Parker, Brenda Spence
Canada East Coast: Hugues Benoit, Daniel Ricard, Kent Smedbol, Hugo Bourdages, Noel Cadigan, David Coffin, Verna Docherty, Monica Finley, David Keith, Jenni McDermid, Joanne Morgan, Mikio Moriyasu, Steve Trottier
Australia: Michael Steer, Stephen Mayfield, Richard McGarvey, Jemery Day, Tony Smith
Japan: Naoaki Kono, Tetsuichiro Funamoto, Ryuji Yukami, Tohya Yasuda, Yasuhiro Kamimura,Yoji Narimatsu, Toshiki Kaga, Kunihiko Fujiwara, Mari Yoda, Shingo Watari, Momoko Ichinokawa
Russia: Vladimir Radchenko
South Africa: Johann Augustyn, Richard Ball, Anabela Brandão, Doug Butterworth, Deon Durholtz, Tracey Fairweather, Susan Johnston, Rob Leslie, Genevieve Maharaj, Éva Plagányi-Lloyd, Rebecca Rademeyer, Daniel van Zyl
Morocco: Abdelmalek Faraj
Peru: Renato Guevara Carrasco
Chile: Juan Carlos Quiroz
Norway: Leif Nøttestad
Iceland: Gunnar Stefansson
Faroe Islands: Petur Steingrund, Luis Ridao
United Kingdom: Nicola Walker, Helen Dobby, Coby Needle, Simon Jennings
Ireland: Colm Lordan, Dave Stokes, Hans Gerritsen, Maurice Clarke
Spain: Dorleta Garcia
Denmark: Søren Anker Pedersen, Anne Mette, Margit Eero
Netherlands: Thomas Brunel, Jan Jaap Poos
Poland: Jan Horbowy
Latvia: Georgs Kornilovs
Finland: Jukka Pönni

References appearing only in Supplementary Information:

41. Pella, J. J. & Tomlinson, P. K. A generalized stock production model. *Bull. I-ATTC* **13**, 416–469 (1969).
42. Thorson, J. T., Cope, J. M., Branch, T. A. & Jensen, O. P. Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. *Can. J. Fish. Aquat. Sci.* **69**, 1556–1568 (2012).
43. Fox, W. W. An exponential surplus-yield model for optimizing exploited fish populations. *Trans. Am. Fish. Soc.* **99**, 80–88 (1970).
44. Froese, R. & Pauly, D. FishBase. www.fishbase.org, version (08/2019). World Wide Web electronic publication. www.fishbase.org (2019).
45. Palomares, M. L. D. & Pauly, D. SeaLifeBase. www.sealifebase.org, version (08/2019). World Wide Web electronic publication. (2019).
46. Hyndman, R. *et al. forecast: Forecasting functions for time series and linear models*. (2019).
47. Mace, P. M., Sullivan, K. J. & Cryer, M. The evolution of New Zealand's fisheries science and management systems under ITQs. *ICES J. Mar. Sci.* **71**, 204–215 (2014).
48. Smith, A. D. M. *et al.* Implementing harvest strategies in Australia: 5 years on. *ICES J. Mar. Sci.* **71**, 195–203 (2014).
49. Ichinokawa, M., Okamura, H., Watanabe, C., Kawabata, A. & Oozeki, Y. Effective time closures: quantifying the conservation benefits of input control for the Pacific chub mackerel fishery. *Ecol. Appl.* **25**, 1566–1584 (2015).
50. Gullestad, P. *et al.* Changing attitudes 1970–2012: evolution of the Norwegian management framework to prevent overfishing and to secure long-term sustainability. *ICES J. Mar. Sci.* **71**, 173–182 (2013).
51. Arnason, R. On the ITQ fisheries management system in Iceland. *Rev. Fish Biol. Fisher.* **6**, 63–90 (1996).
52. Jacobsen, J. Path dependence in Faroese fisheries (mis)management. *Mar. Policy* **108**, 103615 (2019).
53. Lassen, H., Kelly, C. & Sissenwine, M. ICES advisory framework 1977–2012: From Fmax to precautionary approach and beyond. *ICES J. Mar. Sci.* **71**, 166–172 (2013).
54. Shumway, R. H. & Stoffer, D. S. *Time Series Analysis and its Applications*. (Springer, 2017).
55. FAO. *Fishery and Aquaculture Statistics: Global capture production 1950-2017* (FishstatJ). Food and Agriculture Organization of the United Nations, Fisheries and Aquaculture Department, Rome (2019).
56. Kitanidis, P. K. & Bras, R. L. Real-time forecasting with a conceptual hydrologic model: 2. Applications and results. *Water Resour. Res.* **16**, 1034–1044 (1980).

Supplementary Table 3 | Regional representation of stocks included in analysis compared to other datasets. Summed catches are means of individual stocks over the period 1970-2017, summed across stocks in the region.

	This analysis		RAMLDB[a]				FAO landings[b]	
Country/region	Number of stocks	Summed catch across stocks (t)	Number of stocks	% incl[c]	Summed catch across stocks (t)	% incl[c]	Summed catch in region (t)	% incl[c]
Australia	16	20,866	25	64%	52,811	40%	169,694	12%
Canada-East Coast	11	282,055	45	24%	824,856	34%	927,900	30%
Canada-West Coast	17	39,579	31	55%	47,174	84%	200,862	20%
Europe(EU) NE Atl	18	2,533,075	88	20%	5,666,003	45%	4,914,943	52%
Europe(non-EU) NE Atl	10	2,780,553	20	50%	3,831,029	73%	5,078,099	55%
Europe-Med/Black Sea	20	393,230	80	25%	577,449	68%	1,488,293	26%
Japan	24	3,636,297	37	65%	3,894,633	93%	5,842,412	62%
New Zealand	23	128,859	49	47%	232,937	55%	354,370	36%
Russia-East Coast	3	1,702,043	3	100%	1,702,043	100%	2,839,219	60%
South Africa	13	443,259	15	87%	553,267	80%	800,467	55%
South America	12	6,690,661	34	35%	9,141,082	73%	10,549,002	63%
US-Alaska	22	1,810,647	43	51%	2,015,489	90%	1,804,418	100%
US-Northeast	31	642,714	39	79%	710,751	90%	1,134,541	57%
US-Southeast	21	79,496	40	53%	705,204	11%	971,843	8%
US-West Coast	20	343,354	51	39%	397,692	86%	430,498	80%
West Africa	5	1,035,381	5	100%	1,035,381	100%	1,545,305	67%
Tuna RFMOs	22	1,269,592	38	58%	3,192,889	40%	5,224,159	24%
Other			1	0%	1,251	0%	35,276,664	0%
Total	**288**	**23,831,661**	**644**	**45%**	**34,581,943**	**69%**	**79,552,688**	**30%**

[a] Includes all available stocks in the RAM Legacy Stock Assessment Database[22] for which a time series of U/U_{REF} and/or B/B_{REF} is available. U_{REF} and B_{REF} may be extracted directly from a stock assessment, or may be estimated *post-hoc* using a surplus production model, as described above.

[b] Extracted from the Global Capture Production database of the Food and Agriculture Organization of the United Nations[55]. FAO Major Fishing Areas do not align exactly with the regions considered in this analysis, so regional catch totals are approximate. Regional sums are calculated based on the relevant country(ies) and on inclusions of the following FAO Major Fishing Areas:

Australia (Ind-E-57, Pac-WC-71, Pac-SW-81)
Canada-East Coast (Atl-NW-21)
Canada-West Coast (Pac-NE-67)
Europe(EU) NE Atl (Atl-NE-27 for EU countries)
Europe(non-EU) NE Atl (Atl-NE-27 for Norway, Iceland, Faroe Islands, Russia)
Europe-Med/Black Sea (Med-37)
Japan (Pac-NW-61)
New Zealand (Pac-SW-81)
Russia-East Coast (Pac-NW-61)
South Africa (Atl-SE-47, Ind-W-51)
South America (Atl-SW-41, Pac-SE-87 for Peru, Chile, Argentina)
US-Alaska (87.2% of U.S. landings in Pac-NE-67; proportion based on NOAA catch-by-state landings data)
US-Northeast (Atl-NW-21)
US-Southeast (Atl-WC-31)
US-West Coast (Pac-EC-77, and 12.8% of U.S. landings in Pac-NE-67; proportion based on NOAA catch-by-state landings data)
West Africa (Atl-EC-34 for Morocco, Mauritania, Senegal, Gambia, Guinea-Bissau, Guinea, Sierra Leone, Liberia, Cabo Verde)

All of the above summed FAO landings omitted freshwater ISSCAAP taxonomic groups, and also omitted the ISSCAAP group 'tunas, bonitos, billfishes' to avoid double-counting. For the region 'Tuna RFMOs', landings from this ISSCAAP group 'tunas, bonitos, billfishes' were summed over all countries and FAO Major Fishing Areas. For the remaining region 'Other', the sum of regional sums was subtracted from the summed mean global marine landings to represent the portion from regions in which formal stock assessments are less commonly conducted.

[c] Percentages included (% incl) are the number of stocks (or summed catch) in this analysis as a proportion of the number of stocks (or summed catch) from all available stocks (or summed catch) in RAMLDB or the FAO landings database.

Supplementary Table 4 | Levels of relative fishing pressure (U/U_{REF}) and relative biomass (B/B_{REF}) at distinct points in the time series of stocks included in analyses. Years at which the condition applied are also summarised. Summarised values for years, U/U_{REF}, and B/B_{REF} include the number of stocks (n) and percentiles of the distribution of values across stocks, as well as the proportion of stocks depleted below the level of $B/B_{\mathrm{REF}} = 0.5$.

	Year[a]				U/U_{REF}				B/B_{REF}				
Point in time series	n[b]	25th %ile	50th %ile	75th %ile	n	25th %ile	50th %ile	75th %ile	n	25th %ile	50th %ile	75th %ile	% stocks < 0.5[c]
Year 1950	67				65	0.09	0.28	0.60	66	1.58	2.16	2.83	5%
Start of full time series	288	1952	1973	1983	286	0.21	0.66	1.48	282	0.84	1.60	2.36	12%
Start of mature fishery phase	288	1963	1977	1986	286	0.49	0.87	1.53	282	0.82	1.57	2.30	12%
First year under rebuilding plan	118	1992	1999	2005	115	0.51	1.40	2.50	118	0.22	0.40	0.64	64%
First year of scientific surveys	237	1979	1985	1996	223	0.66	1.18	2.06	226	0.50	0.99	1.57	26%
First year of stock assessment	277	1985	1995	2002	264	0.63	1.30	2.36	269	0.44	0.89	1.57	30%
First year of harvest control rule	194	1990	1998	2005	184	0.65	1.10	1.80	192	0.41	0.76	1.36	34%
First year of fleet-wide catch limits	221	1983	1992	1998	210	0.62	1.15	2.01	217	0.52	0.99	1.63	24%
First year of individual quotas	139	1988	1997	2003	133	0.53	1.10	1.93	136	0.50	0.90	1.67	25%
Year of EEZ declaration	253	1976	1978	1982	237	0.48	1.02	2.06	245	0.58	1.25	2.08	22%
Year of UN CA/FSA ratification	278	1995	1996	2000	268	0.71	1.27	2.03	271	0.42	0.81	1.35	30%
Year of national/ regional policy	239	1988	1996	1996	229	0.51	1.07	1.99	238	0.48	0.92	1.64	27%

[a]Years of first use of management measures are constrained to the range of years for which time series of U/U_{REF} and/or B/B_{REF} were available, so that these years correspond with the values of U/U_{REF} and/or B/B_{REF} in columns further to the right. If a management measure was implemented for a stock before its first available values of U/U_{REF} or B/B_{REF}, for summary purposes in this table its year of first use is considered to be the first year of available U/U_{REF} or B/B_{REF}.

[b]All 288 stocks included in analyses had a mature fishery phase, and therefore also had a full time series (including the developing phase). Relatively few stocks (67) had a time series of U/U_{REF} and/or B/B_{REF} extending back to 1950. Some stocks have never had a given management measure applied, therefore sample sizes associated with first use of the measure are less than 288.

[c]Fraction of stocks with $B/B_{\mathrm{REF}} < 0.5$ at the distinct point in the time series listed. We note that a variety of thresholds around the world are used to define ‘overfished’ or ‘depleted’; the value of 0.5 considered here and shown for consistency is a common threshold, but by no means the only one. Following footnote (a), if a management measure was implemented for a stock before its first available value of B/B_{REF}, its first value of B/B_{REF} in the time series was compared relative to 0.5 in the calculated fraction across stocks.

Supplementary Table 5 | Required order of differencing of response variable time series to ensure stationarity[a].

		Differences required		
Fishery phase	Response variable	0	1	2
Mature	$\ln(U/U_{\mathrm{REF}})$	96	166	22
Full time series	$\ln(U/U_{\mathrm{REF}})$	81	172	31
Mature	$\ln(B/B_{\mathrm{REF}})$	52	156	72
Full time series	$\ln(B/B_{\mathrm{REF}})$	47	163	70

[a] Values reflect frequencies of stocks for which the order of differencing was sufficient based on a one-sided KPSS test for stationarity, implemented with the auto.arima() function of the R package ‘forecast’[46]. Time series for each stock were separated into ‘developing’ (not of interest for analyses) and ‘mature’ (of interest) phases, and the full time series was also evaluated. A minimum of 10 years of data per stock per phase were required for evaluation of a given response variable.

Supplementary Table 6 | Best-fit orders of autoregressive (*p*) and moving average (*q*) components of response variable time series to maximise goodness-of-fit[a].

		$\ln(U/U_{\mathrm{REF}})$						$\ln(B/B_{\mathrm{REF}})$					
		$q \rightarrow$						$q \rightarrow$					
Fishery phase	*p*	0	1	2	3	4	5	0	1	2	3	4	5
Mature	0	123	47	19	9	5	0	41	44	12	8	4	3
	1	28	4	0	2	1	0	65	18	9	1	0	0
	2	22	2	1	1	0	0	33	6	1	0	0	0
	3	7	2	1	0	0	0	13	2	0	0	0	0
	4	3	1	0	0	0	0	10	1	0	0	0	0
	5	3	0	0	0	0	0	7	0	0	0	0	0
Full time series	0	115	42	18	14	4	0	41	39	13	10	6	2
	1	31	5	1	3	2	0	65	16	10	1	2	0
	2	22	3	0	1	0	0	33	8	0	1	0	0
	3	10	3	2	0	0	0	11	3	0	0	0	0
	4	3	1	0	0	0	0	9	1	0	0	0	0
	5	4	0	0	0	0	0	9	0	0	0	0	0

[a] Values reflect frequencies of stocks for which the combined order of *p* and *q* minimised the AICc for first-order differenced time series, implemented with the auto.arima() function of the R package 'forecast'[46]. Time series for each stock were separated into 'developing' and 'mature' phases, and the full time series was also evaluated. A minimum of 10 years of data per stock per phase were required for evaluation of a given response variable.

Supplementary Table 7 | Model fit diagnostics for ARIMA models fit to response variables $\Delta\ln(U/U_{REF})$ and $\Delta\ln(B/B_{REF})$. Three metrics are summarised for each of two ARIMA model structures from different sections of the analysis. Summaries include the number of stocks (n), percentiles of the distribution of values across individual stocks, as well as an overall value of the metric across all stocks. .

		$\Delta\ln(U/U_{REF})$					$\Delta\ln(B/B_{REF})$				
Metric	Results section[a]	n	25th %ile	50th %ile	75th %ile	Overall across stocks[b]	n	25th %ile	50th %ile	75th %ile	Overall across stocks[b]
Mean error[c]	*i*	284	-0.02	0.00	0.02	0.000868	280	-0.02	0.00	0.02	0.000037
	iii	277	-0.04	0.00	0.03	0.001186	277	-0.02	0.00	0.02	0.000001
Root mean square error (RMSE)[d]	*i*	284	0.23	0.32	0.49	0.58	280	0.09	0.14	0.25	0.22
	iii	277	0.24	0.34	0.53	0.59	277	0.08	0.13	0.25	0.21
Persistence index[e]	*i*	284	0.41	0.54	0.62	0.60	280	-1.11	-0.09	0.37	0.42
	iii	277	0.42	0.54	0.62	0.59	277	-0.92	0.02	0.38	0.45

[a] Results section *i* , '*Base model for stock status trends*', corresponds to the base model described in Equation 1, with results presented in Figure 3. Results section *iii*, '*Predicting equilibrium responses to management*' corresponds to the coupled-variable model described in Equation 3, with results presented in Figure 5. Model fit diagnostics are not shown for Results section *ii*, '*Predicting short-term responses to management*', corresponding to Equation 2 (and Figure 4), but are similar to those listed for section *i*.

[b] Statistical fits for the hierarchical model were for all stocks simultaneously rather than for each stock individually, thus these overall values of metrics across all stocks better reflect model performance than the percentiles of the distributions of individual stocks.

[c]Mean error is the mean of predicted values minus observed values, either for an individual stock's time series or for the combined dataset across all stocks. It is calculated over years *t* as:

$$\frac{1}{T}\sum_{t=1}^{T}\left(y_{pred,t} - y_{obs,t}\right)$$

For comparison with these mean errors, overall means of response variable values across all stocks were: for $\Delta\ln(U/U_{\mathrm{REF}})$ section *i*, observed -0.0098, predicted -0.0107; for $\Delta\ln(U/U_{\mathrm{REF}})$ section *iii*, observed 0.0122, predicted 0.0110; for $\Delta\ln(B/B_{\mathrm{REF}})$ section *i*, observed -0.0127, predicted -0.0127; and for $\Delta\ln(B/B_{\mathrm{REF}})$ section *iii*, observed -0.0120, predicted -0.0120.

[d]RMSE represents the standard deviation of the model prediction error, calculated as the square root of the mean of squared deviations between observed and predicted values. This is calculated either for an individual stock's time series or for the combined dataset across all stocks. Values of RMSE are not meant to be compared between $\Delta\ln(U/U_{\mathrm{REF}})$ and $\Delta\ln(B/B_{\mathrm{REF}})$, nor between sections *i* and *iii*, as these represent different response variables or sample sizes. It is calculated over years *t* as:

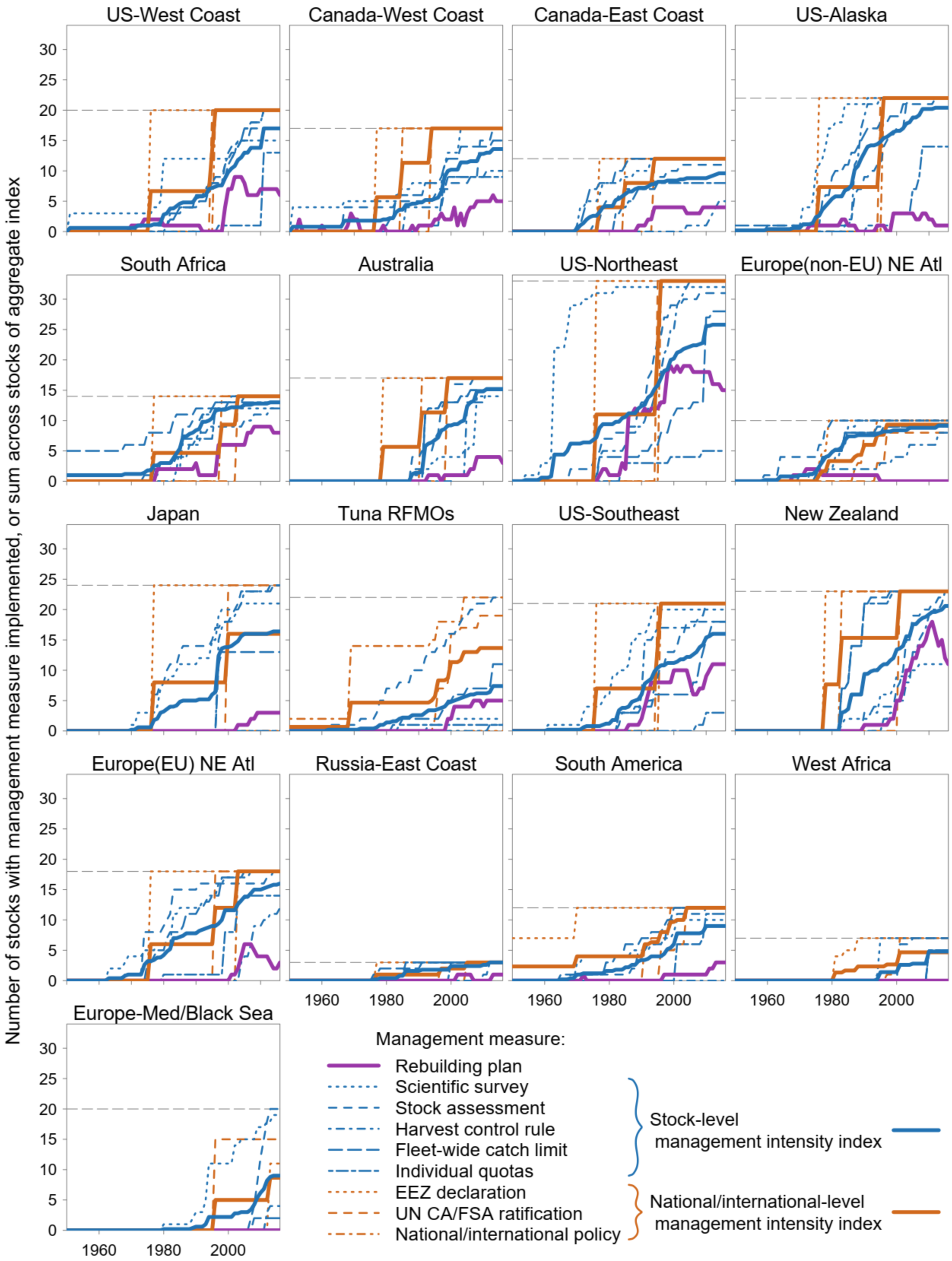
$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(y_{pred,t} - y_{obs,t}\right)^2}$$

[e]Persistence index, or coefficient of persistence[56], compares model performance (predicted versus observed values) against interannual changes in the observed values (from one year to the next). The latter changes represent a simpler model in which the observed value from the previous year represents the prediction for the current year. It is calculated either for an individual stock's time series or for the combined dataset across all stocks (in which the first year for each stock is omitted), as:

$$1 - \frac{\sum_{t=2}^{T}\left(y_{pred,t} - y_{obs,t}\right)^2}{\sum_{t=2}^{T}\left(y_{obs,t} - y_{obs,t-1}\right)^2}$$

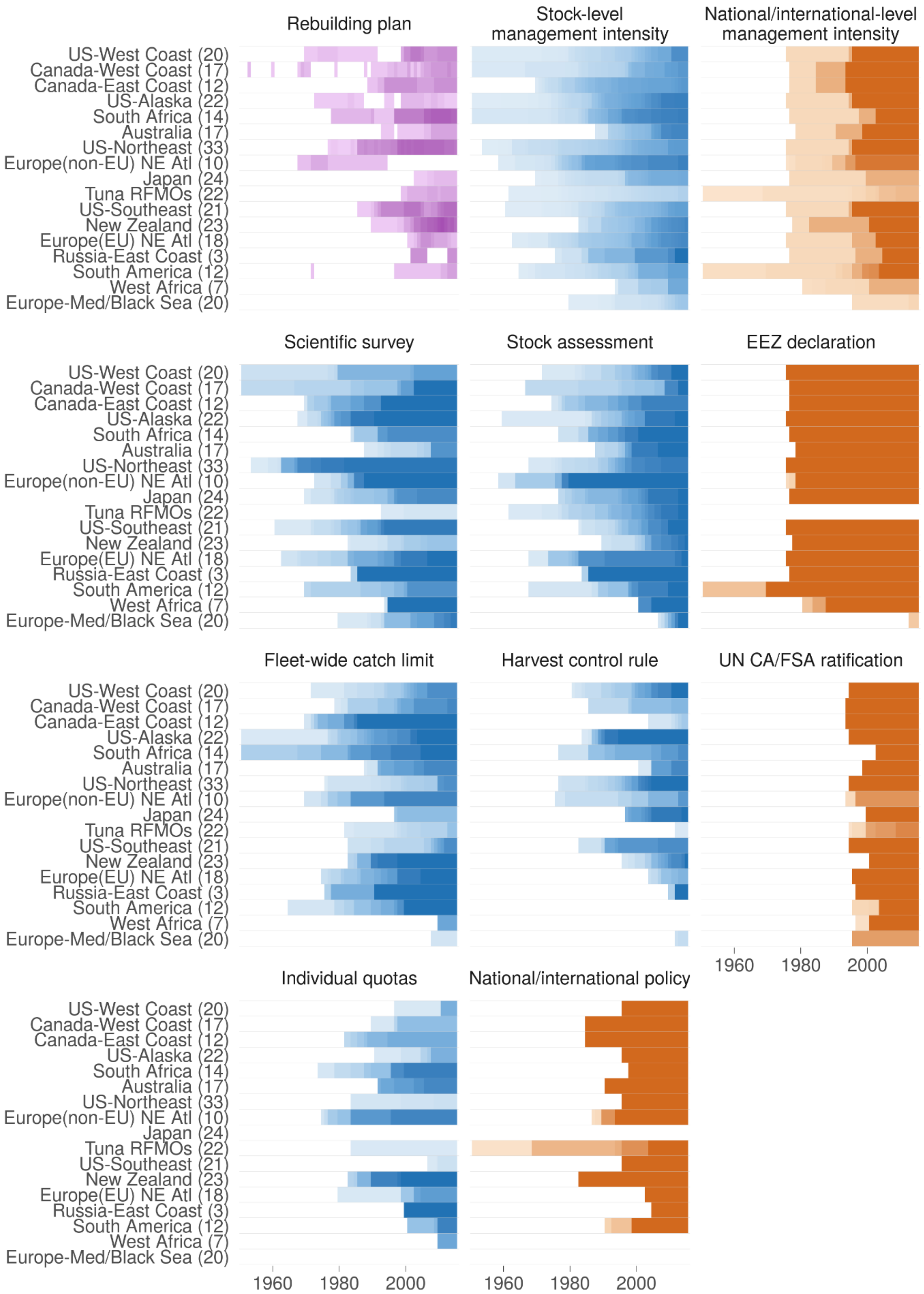
Values typically range from 0-1, with values of 1 reflect perfect model performance, and values ≤0 reflecting poor predictive performance.

Supplementary Figure 2 (next page) | Implementation history of fisheries management measures in 17 regions. Values for individual measures reflect the number of stocks in the region with the measure implemented over time. For rebuilding plans, the thick purple line shows the number of stocks currently under a rebuilding plan in any given year, so counts in the region may decrease as rebuilding plans are de-activated following recovery of a stock. For all other individual measures, counts increment with their first use on a stock but do not revert downwards if the measure is later ceased (e.g., if a survey or an assessment is not conducted in any given year), so counts represent current or previous usage of the measure. Other thick solid lines represent aggregate indices of management intensity at the stock level (blue; comprising the five management measures indicated) or at the national/international level (orange; comprising the three measures indicated). These aggregate indices range from 0-1 for each stock depending on how many of the component measures have been implemented, so lines show sums across stocks in the region and do not revert downwards. Dashed grey horizontal lines show the total number of stocks from the region included in analyses. Regions are ordered left-to-right, top-to-bottom by median U/U_{REF} across stocks over their last five years of available data (lowest-to-highest). Data shown are for the same 288 stocks as shown in Fig. 1a, here separated by region. Management measures are described in Supplementary Table 1 and implementation histories of individual stocks are shown in Supplementary Figure 1. See Supplementary Discussion for further description of this figure.

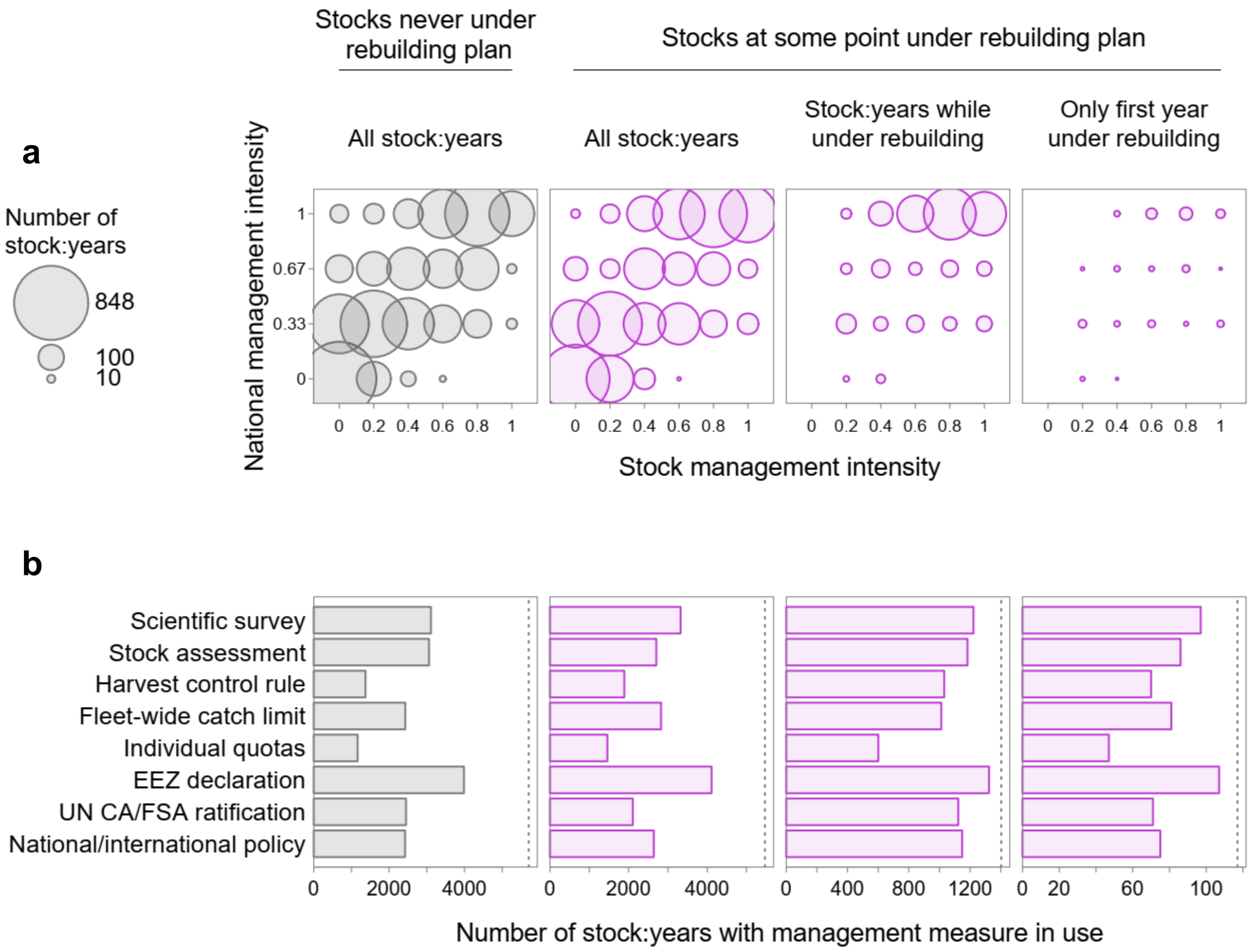


Supplementary Figure 2 continued

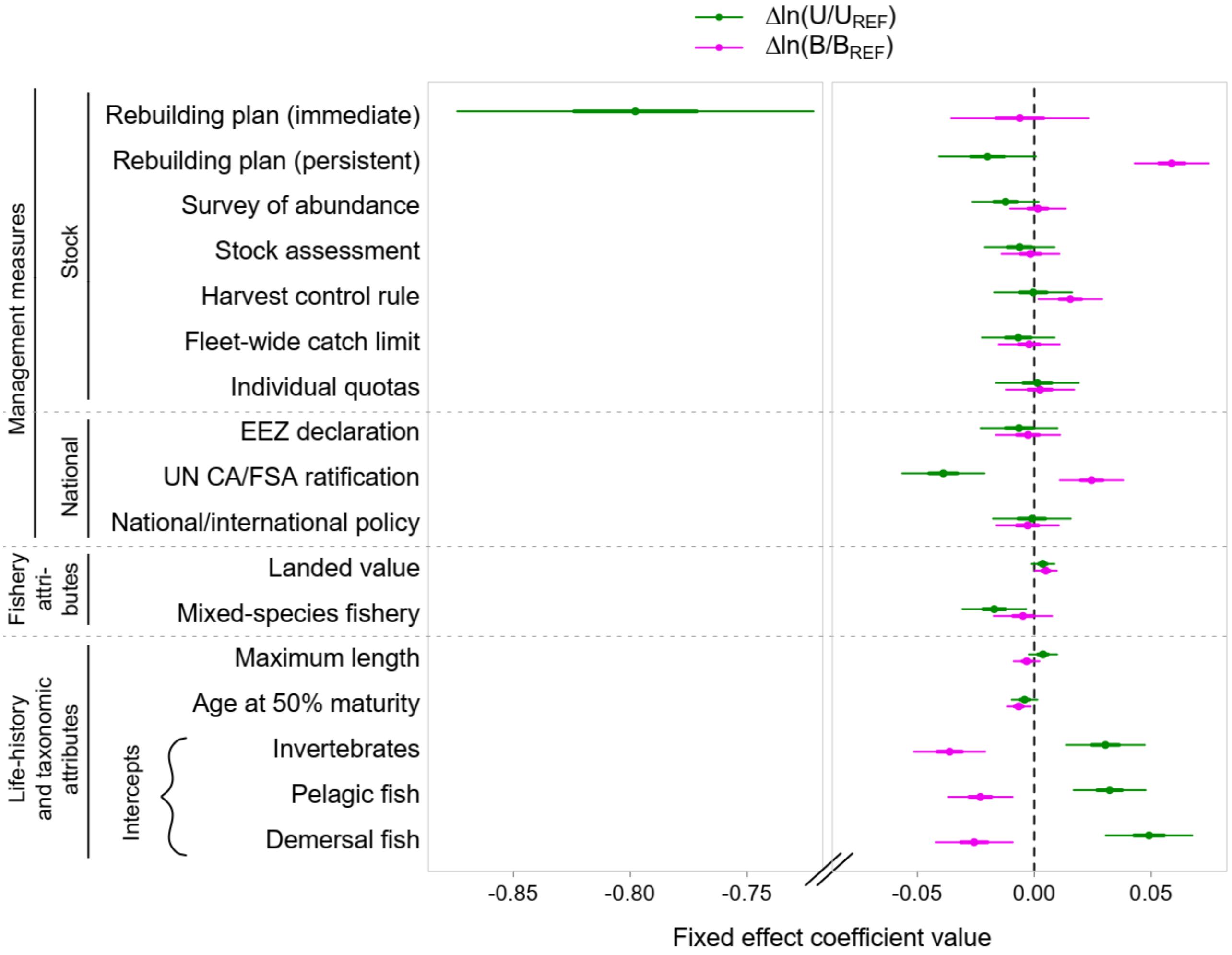
Supplementary Figure 3 (next page) | Implementation history by region of nine fisheries management measures and two aggregate indices. In each panel, shading for each region reflects the proportion of stocks with the measure implemented. For rebuilding plans, proportions show stocks currently under a rebuilding plan in any given year, so proportions may decrease as rebuilding plans are de-activated. For all other individual management measures in the lower eight panels, proportions increment with first use of the measure for a stock, but do not revert downwards if the measure is later ceased. Top row panels for aggregate indices of management intensity at the stock level (comprising five measures) or at the national/international level (comprising three measures) show joint proportions of stocks and index values. The number of stocks in each region is shown in parentheses. Regions are ordered top-to-bottom by median U/U_{REF} across stocks over their last five years of available data (lowest-to-highest). The top three panels were shown in Fig. 1b. Management measures are described in Supplementary Table 1 and implementation histories of individual stocks are shown in Supplementary Figure 1. See Supplementary Discussion for further description of this figure.



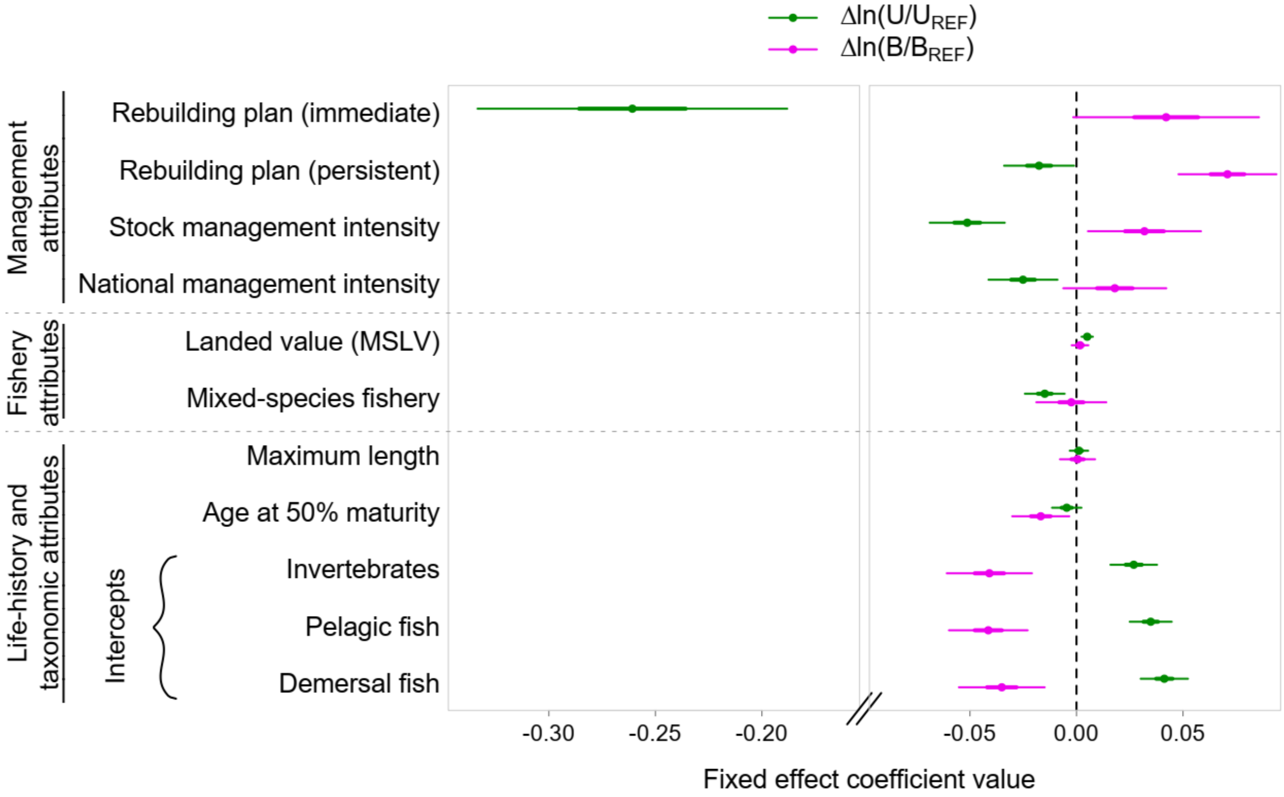
Supplementary Figure 3 continued



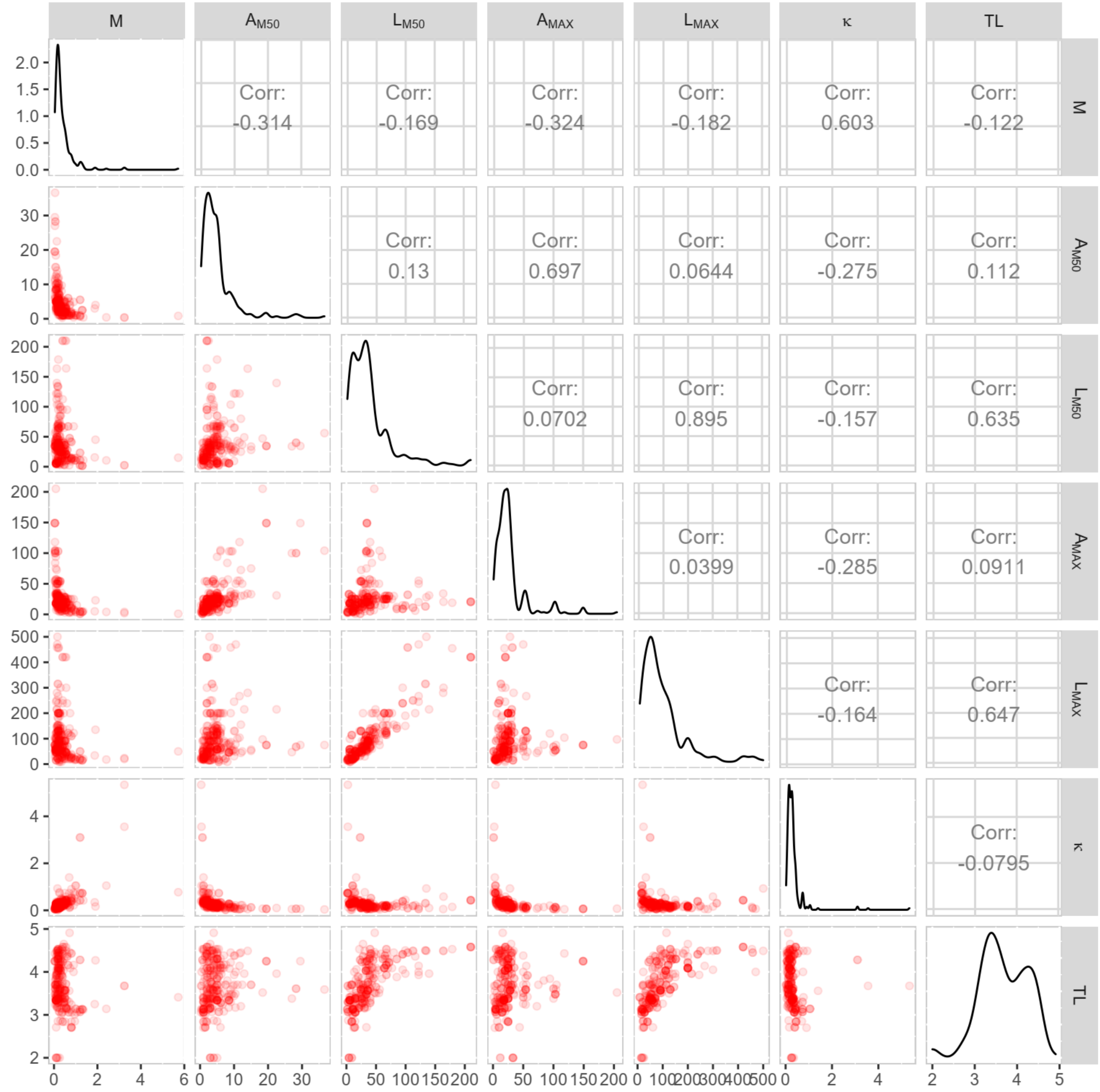
Supplementary Figure 4 | Associations between implementing rebuilding plans and concurrent use of other management measures. Frequencies of management measures in use are shown for four data subsets: all years of stocks that have never been under a rebuilding plan (grey); all years of stocks that have at some point been under a rebuilding plan (purple, left); years while under an active rebuilding plan (purple, middle); and a stock's first year under a rebuilding plan (purple, right). In (**a**), panels show all possible combinations of stock-level management intensity and national/international-level management intensity. Symbol size is proportional to the number of stock:years of data at each combination. In (**b**), panels show frequencies of individual management measures in use. Vertical dotted lines show the total number of stock:years in the data subset.



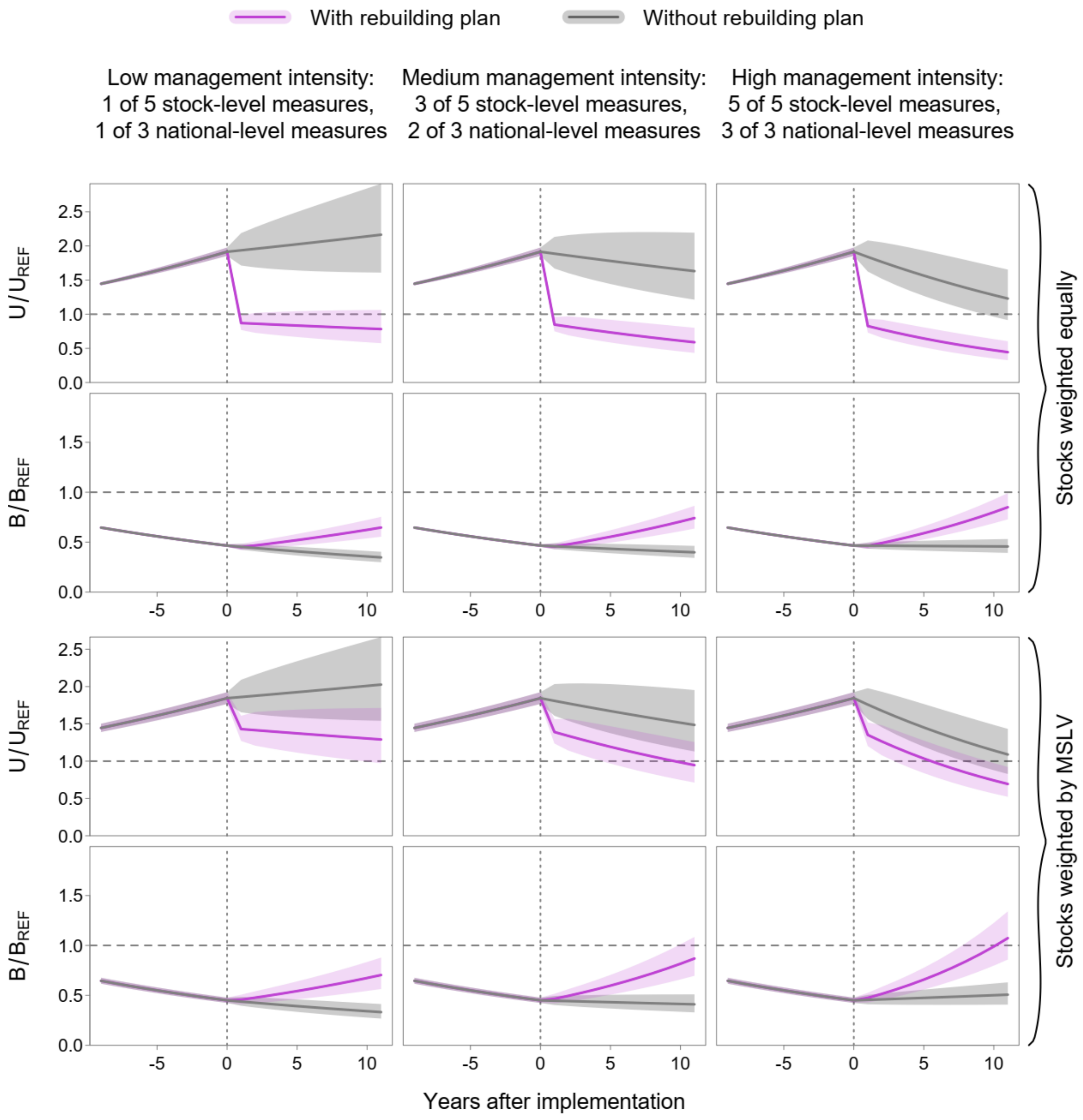
Supplementary Figure 5 | Effects of individual management measures, fishery attributes, and life-history traits on annual changes in relative fishing pressure and relative biomass. Positive (or negative) coefficients reflect increasing (or decreasing) trends in fishing pressure (U/U_{REF}) and biomass (B/B_{REF}) during the 'mature' fishery phase. The horizontal axis is broken for visual clarity, as one coefficient differs substantially in magnitude from the others. Model structure is identical to the base model (main Methods *i* and Fig. 3), except the individual components of management intensity indices (five stock-level components and three national-level components) are included as predictors instead of the two aggregate indices. Stocks are weighted equally. The reference group for overall intercepts is 'Single-species fishery', with the categorical 'Mixed-species fishery' representing a difference from these intercepts. Thick and thin error bars represent standard errors and 95% confidence intervals, respectively.



Supplementary Figure 6 | Effects of management, fishery, and life-history attributes on annual changes in relative fishing pressure and relative biomass under an alternative weighting assumption. Positive (or negative) coefficients reflect increasing (or decreasing) trends in fishing pressure (U/U_{REF}) and biomass (B/B_{REF}) during the ‘mature’ fishery phase. The horizontal axis is broken for visual clarity, as one coefficient differs substantially in magnitude from the others. Model structure is identical to the base model (main Methods *i* and Fig. 3), except instead of equal weighting, stocks are weighted by maximum sustainable landed value (*MSLV*), the product of maximum sustainable yield and average ex-vessel price. The reference group for overall intercepts is ‘Single-species fishery’, with the categorical ‘Mixed-species fishery’ representing a difference from these intercepts. Thick and thin error bars represent standard errors and 95% confidence intervals, respectively.

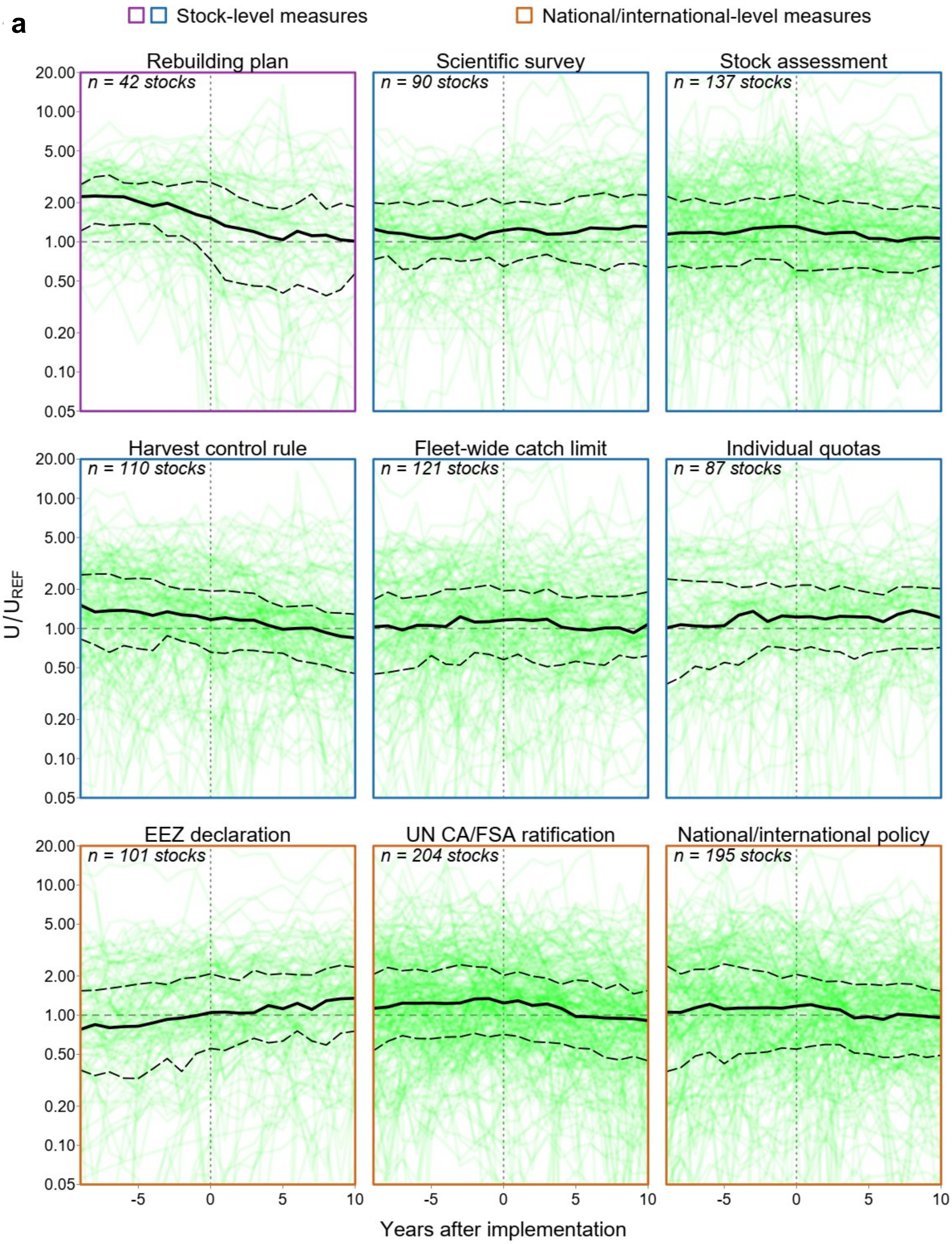


Supplementary Figure 7 | Associations between life-history traits initially considered as potential predictor variables for analyses. Scatterplots (lower panels) and correlation coefficients (upper panels) are shown for of all pairs of variables: natural mortality rate, M; age at 50% maturity, A_{M50}; length at 50% maturity, L_{M50}; maximum age, A_{MAX}; maximum length, L_{MAX}; von Bertalanffy growth, κ; and trophic level, TL. Density plots are shown on the diagonal. As a result of strong correlation with other traits, most traits were omitted as predictor variables. Only A_{M50} and L_{MAX} were carried forward as predictors in regression analyses.

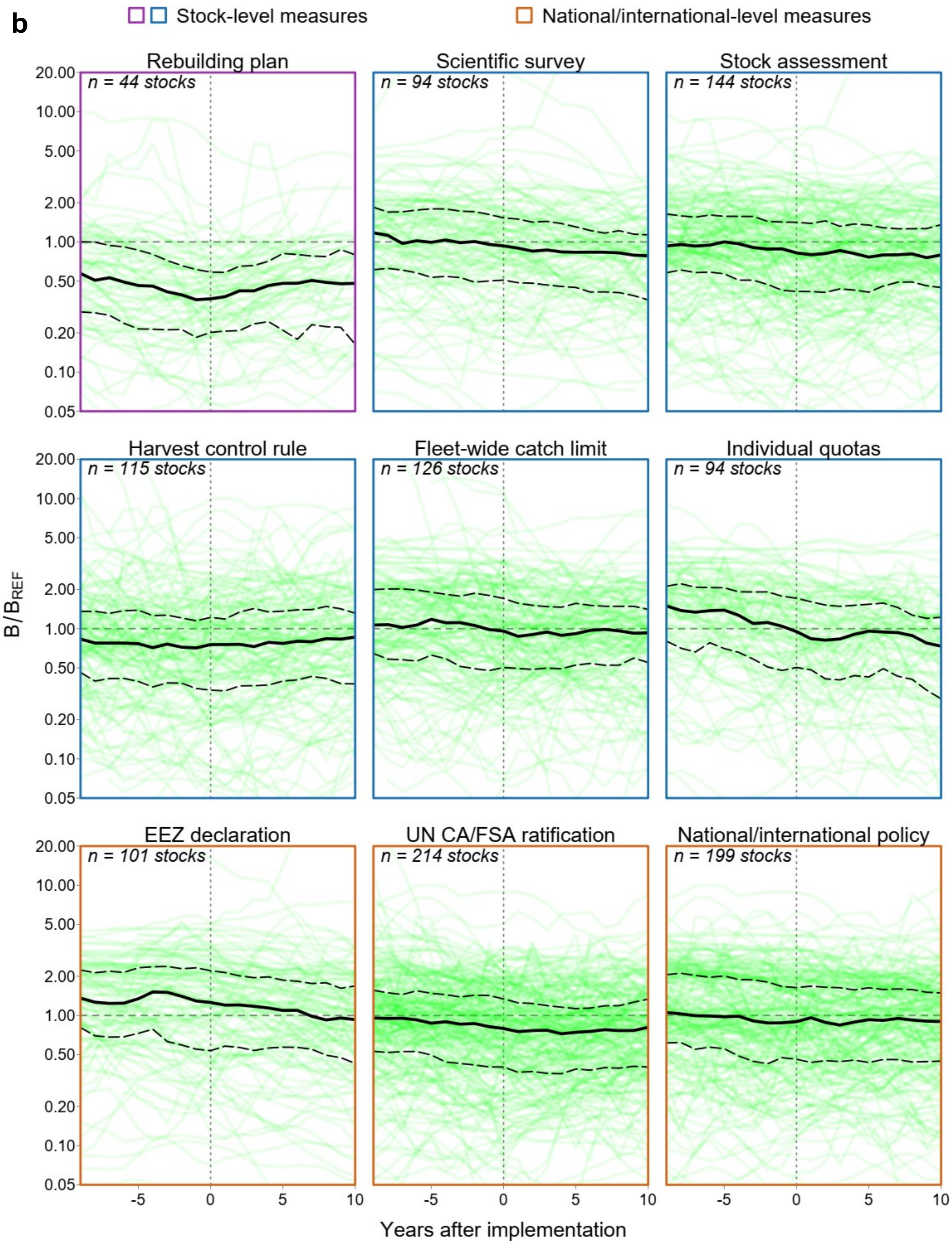


Supplementary Figure 8 | Predicted effects of fisheries management interventions on stock status under varying levels of management intensity and alternative weighting assumptions. Predictions are shown for low (left panels), medium (middle panels), and high (right panels) levels of management intensity. For a given management intensity, the number of measures indicated are implemented in year 0, either with or without a rebuilding plan. Stocks are either equally-weighted (top two rows) or weighted by maximum sustainable landed value (bottom two rows). The top two right-most panels were shown in Fig. 4; see that caption for further details.

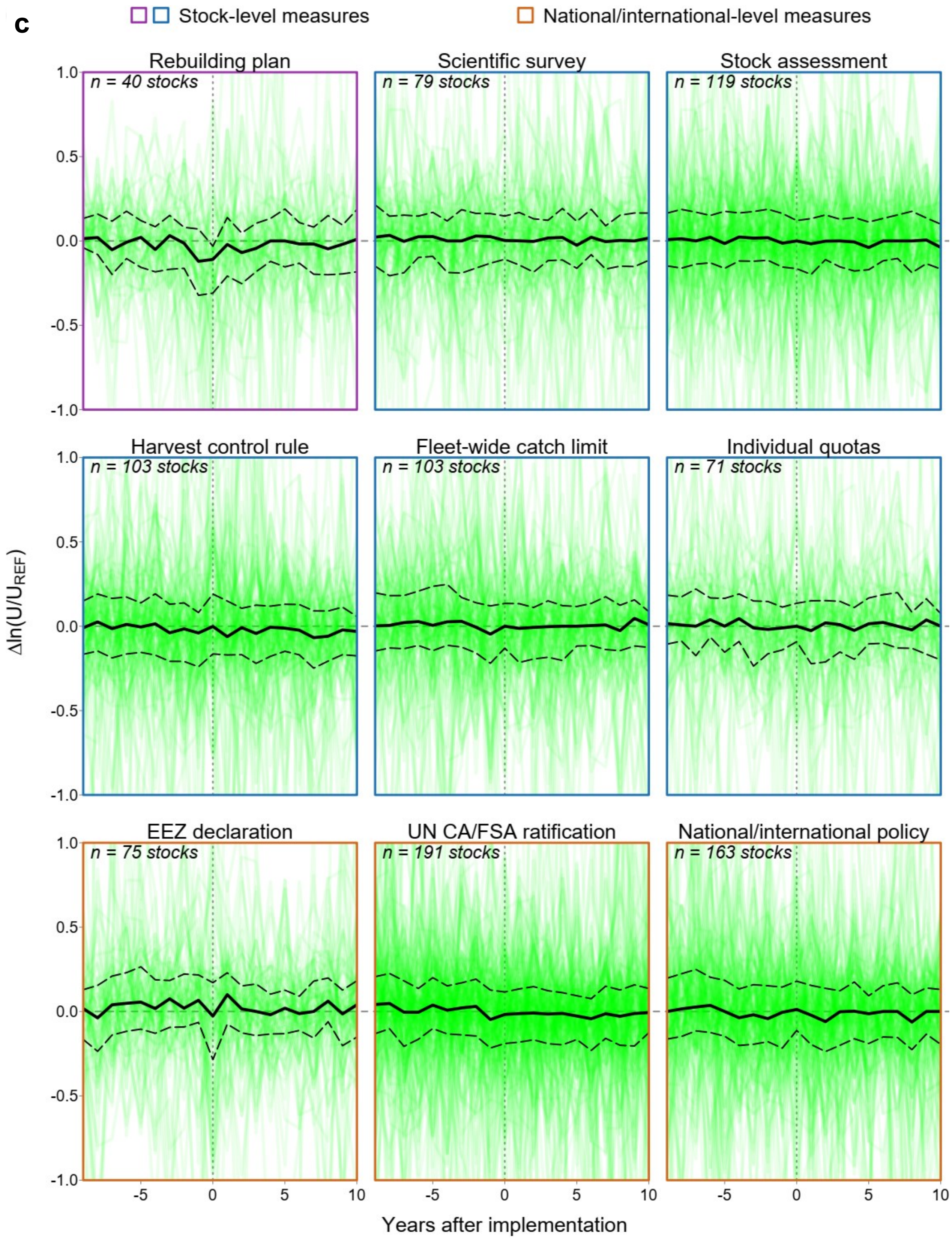
Supplementary Figure 9 (next four pages) | Time series of relative fishing pressure, relative biomass, and annual changes in these variables before and after implementation of management measures. Values of ratios (**a**) U/U_{REF}, and (**b**) B/B_{REF}, are shown on log scale. Values of annual changes in log-ratios (**c**) $\Delta\ln(U/U_{\mathrm{REF}})$, and (**d**) $\Delta\ln(B/B_{\mathrm{REF}})$, are shown on linear scale. The time series of each stock is shifted horizontally so the measure's implementation coincides with year 0. For (**c**) and (**d**), $\Delta\ln(U/U_{\mathrm{REF}})$ and $\Delta\ln(B/B_{\mathrm{REF}})$ shown at year 0 correspond to the change from year 0 to year 1. Individual measures are the same as those listed in Fig. 1 and Supplementary Table 1. For each measure, a minimum of five years of available data before year 0 and five years after year 0 were required for plotting. For rebuilding plans, year 0 represents the first year of at least five consecutive years under rebuilding and follows at least five consecutive years that were not under a rebuilding plan. Sample sizes indicated for each attribute are the number of stocks meeting these plotting requirements; sample sizes for (**c**) and (**d**) are lower than those for (**a**) and (**b**) because a minimum of six years of values are required to calculate the minimum five annual changes in values before and after year 0. The thick solid black line shows the median across stocks with available data in any given year, and thin dashed black lines show 25th and 75th percentiles. Horizontal dashed grey lines show general management targets in (**a**) and (**b**), and show the point of no annual change in the ratios in (**c**) and (**d**).



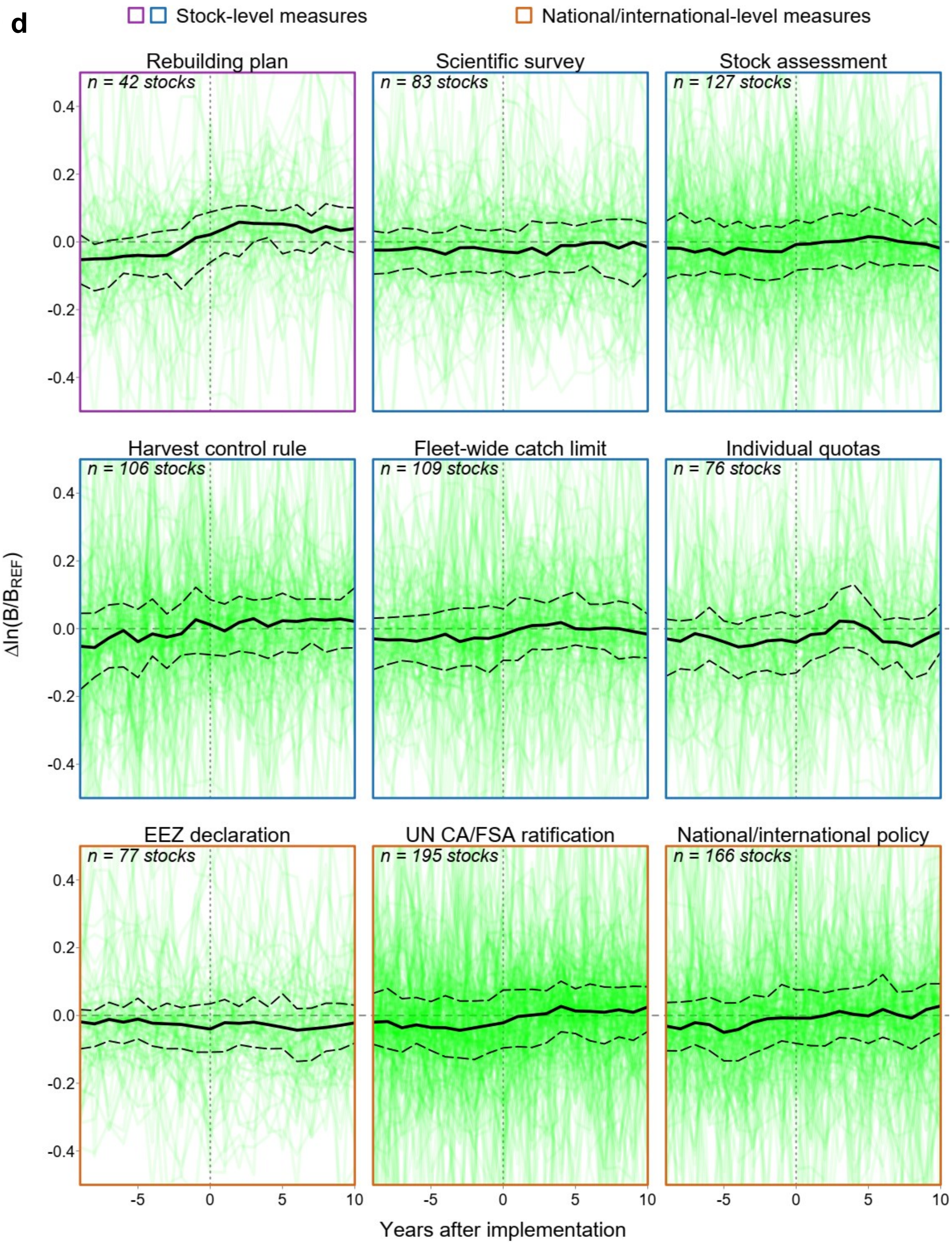
Supplementary Figure 9 continued



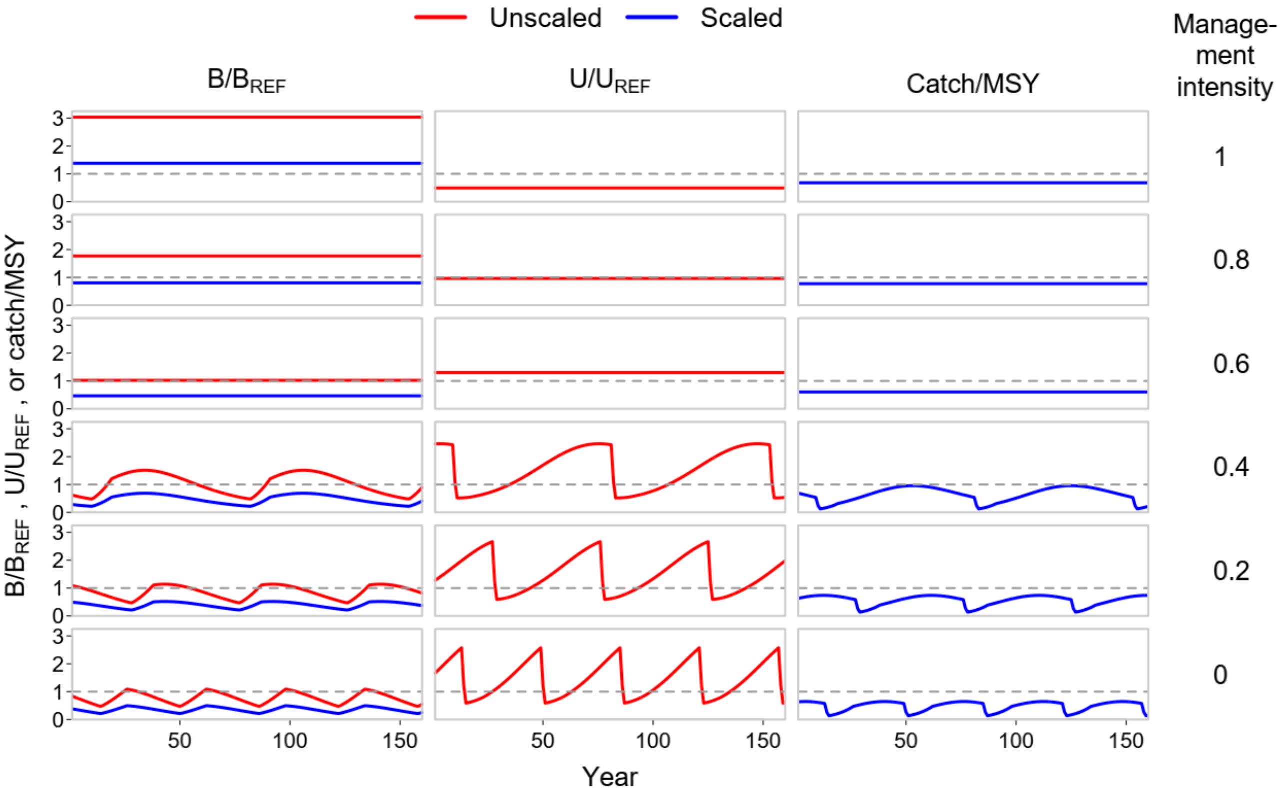
Supplementary Figure 9 continued



Supplementary Figure 9 continued



Supplementary Figure 9 continued



Supplementary Figure 10 | Time series projections at equilibrium under different levels of management intensity. Representative time series of relative fishing pressure (U/U_{REF}), relative biomass (B/B_{REF}, both unscaled and scaled values), and relative catch (catch/MSY, based on scaled biomass) are shown for an average stock. Example values assumed for management intensity apply to both stock-level and national-level indices at the combinations indicated by '×' in Fig. 5. At higher values of management intensity, rebuilding plans were never implemented, and equilibria were stable points. At lower values of management intensity, equilibria switched to stable cycles as rebuilding plans activated in response to low biomass and were then de-activated following stock recovery.