NETHERLANDS INSTITUTE OF ECOLOGY

Scientific use of biodiversity databases

Peter M.J. Herman and Mark J. Costello

Outline

- The big question
- Derived scientific questions
- Data needed
- Data availability
- Data types and suitability
- Data treatment / products
- Data and product governance

The big question

- Indices that indicate the status and the future evolution (further degradation/improvement) of marine ecosystems in Europe
- Assist in managing the future of the sea
- Using current databases, monitoring programmes and scientific knowledge

- Note: concentrate on variables at the community and ecosystem level, not at population level, single species, genes,..

Derived questions. 1. Theoretical

- Can indices be formulated that
  - Relate to ecological theory
  - Relate to both structure and function of the system
  - Are expected to change under anthropogenic pressure
  - Are changing predictably under pressure
- Potential indicators:
  - Species richness (αβγ diversity) in areas (~ sample size, area sampled,..)
  - Trophic structure (levels, productivity, top predators, dominant species, cascades,…)
  - Eutrophication indicators
  - Spatial structure of communities

Derived questions. 2. descriptive

How do biodiversity and ecosystem structure relate to the known environment.

- Which environmental variables are driving at which scales?
  - What environmental maps are needed?
  - What is feasible?
  - What is missing?
  - What can be constructed from indirect evidence?
  - At what scales can the environmental maps be made?

- Which biotic variables are best summarizing the state of the ecosystem
  - Observable and observed
  - Statistically different between different environments
  - Significantly related to present (and hopefully future) pressures (e.g. from comparison with no-take zones)
  - At different and appropriate scales

Derived questions. 3. Statistical

- What is the background variability of indicators in space and time?
- What lumping is needed to describe ‘climate’ and not ‘weather’ – implications for the scales for prediction?
- How can sampling and monitoring be optimized?
- Can useful regression and classification models relate species to habitats? Can these be validated?
- Is prediction of biotic changes possible *via* prediction of changes in habitats?
- Can all this be done at different scales?
- How do alpha-beta-gamma diversity relate spatially?
- How connected are communities (species versus community ranges)?

Data needed

- Species presence, abundance, biomass data *at different scales*
- Species attributes: functional (e.g. trophic role), structural (e.g. reef-forming), tolerance (e.g. temp. range).
- Environmental data *at different (matching) scales*: physical, biogeochemical
- Special attention for biological structural elements (e.g. Reefs, biogenic habitats)
- Gradients (e.g. slopes, fluxes, fronts) or thresholds (tipping points) maybe more important than averages
- Integrity of *habitat networks* (for migrating species) deserves attention

Data availability

- Publish data in a citable way
- Streamline with existing citation scores
- Need for ‘data e-journal’ or collaboration with existing journal
- Archive published data ‘as such’ *(author)+* integrate in larger databases *(data centre)*
- Quality control through peer review? Through data centre?

Data types and suitability

Ecosystem component	Variable	Derived product	Scale
Environment	Temperature, salinity Bathymetry Area/vertical structure	Water masses stratification	Decadal for regional seas Seasonal for high-intensity areas Temp. from RS
	Seabed topography Sediment composition	Benthic habitat (slopes, etc.)	Decadal at most (reg. seas) Semi-decadal (estuaries) RS (intertidal) – seasonal
	Hydrography (currents, stratification)	Bentho-pelagic habitat	Decadal (observations) Seasonal (models)
Biogeochemistry	Nutrients, DOC, POC, SPM, oxygen, pH, alkalinity, pollutants	Potential productivity Pollution status Eutrophication $CaCO_3$ stability Biological pump	Decadal for regional seas Seasonal (coasts, estuaries)

Data types and suitability

Ecosystem component	Variable	Derived product	Scale
Phytoplankton	Chlorophyll, Species composition (HABs)	Productivity Eutrophication status Major foodweb structure Benthic loading	Seasonal (RS ocean colour; automatic monitoring; coastal areas) Yearly to decadal (reg. seas)
Zooplankton	CPR transects, point samples	Population status	Yearly to decadal (reg. seas) Seasonal (some coasts)
Benthos	Density, biomass, species composition	Biotope cover, functional role	Decadal at most (reg. seas); yearly (coasts/est.)
Mobile fauna	Video Trawl?	Abundance	Yearly to decadal
Key species	Abundance Distribution	Population status	Yearly to decadal, mostly coastal/est.

Data types and suitability

- For regional seas (e.g. North Sea) a decadal time scale and 10 km spatial scale is probably maximum attainable
- Coasts and estuaries much better monitored -> apply multi-scale approach
- Use of models and RS needed for synoptic pictures -> integrate in data products?
- Models needed to interpolate between point data (need both)
- Data products must aim at temporal scales justified by the data

Data treatment

- **Standards are necessary for data management**
  - Species names – standard classification available in ERMS (Aphia database)
  - Biotopes – standard classification available at JNCC (BioMar classification) as used by EUNIS
  - Units – need built-in conversions sensitive to accurate number of decimal places (so not to give false sense of accuracy)
  - Chemical variables – reflect comparability between methods of sample treatment and analysis
  - Terminology (vocabulary) – need self-sustaining, peer-reviewed, glossary or dictionary to help streamline use of terminology and explain categories

How to access and manage environmental data?

- Typical physical data / model runs are heavy 4D maps
- Visualisation and exchange tools – OpenDAP
- Based on NetCDF and HDF files, originating from oceanography
- Dutch initiative 'Open Earth' – includes data visualisation and manipulation tools + model output
- Problem: how to integrate with biological data?

Towards an ecosystem product

- Structure environmental data: habitat classifications, oceanographic data, model outputs, remote sensing,…
- Bridge the gap with physical oceanographers!
- Link with biological data sets (species occurrence, density etc.)
- Fill the biogeochemistry gap!
- Establish and validate species-environment relations, estimate species ranges with covariables
- Derive ecosystem-related indicators

Governance issues

- Who 'owns' secondary data?
  - Do creators get special data services or privileged access?
  - Who provides long-term support?
- Is there a role for All Taxon Biodiversity Inventory ATBI Sites as baseline for relative abundance in natural vs. impacted sites
- Compile national, regional inventories and species range mapping?
- How to improve species identification and attribute database?