Build data input for VAST model

Description

`make_data` builds a tagged list of data inputs used by TMB for running the model

Usage

```
make_data(b_i, a_i, t_iz, c_iz = rep(0, length(b_i)), e_i = c_iz[, 1],
  v_i = rep(0, length(b_i)), FieldConfig, spatial_list,
  ObsModel_ez = c(PosDist = 1, Link = 0), OverdispersionConfig = c(eta1
  = 0, eta2 = 0), RhoConfig = c(Beta1 = 0, Beta2 = 0, Epsilon1 = 0,
  Epsilon2 = 0), VamConfig = c(Method = 0, Rank = 0, Timing = 0),
  Aniso = TRUE, PredTF_i = rep(0, length(b_i)), Xconfig_zcp = NULL,
  covariate_data = NULL, formula = ~0, Q_ik = NULL,
  Network_sz = NULL, F_ct = NULL, F_init = 1, t_yz = NULL,
  CheckForErrors = TRUE, yearbounds_zz = NULL, Options = c(),
  Expansion_cz = NULL, Z_gm = NULL,
  Version = FishStatsUtils::get_latest_version(package = "VAST"), ...)
```

Arguments

`b_i`	Sampled biomass for each observation i
`a_i`	Sampled area for each observation i
`t_iz`	Matrix where each row species the time for each observation i (if t_iz is a vector, it is coerced to a matrix with one column; if it is a matrix with two or more columns, it specifies multiple times for each observation, e.g., both year and season)
`c_iz`	Category (e.g., species, length-bin) for each observation i
`e_i`	Error distribution for each observation i (by default `e_i=c_i`)
`v_i`	sampling category (e.g., vessel or tow) associated with overdispersed variation for each observation i (by default `v_i=0` for all samples, which will not affect things given the default values for `OverdispersionConfig`)
`FieldConfig`	a vector of format c("Omega1"=0, "Epsilon1"=10, "Omega2"="AR1", "Epsilon2"=10), where Omega refers to spatial variation, Epsilon refers to spatio-temporal variation, Omega1 refers to variation in encounter probability, and Omega2 refers to variation in positive catch rates, where 0 is off, "AR1" is an AR1 process, and >0 is the number of elements in a factor-analysis covariance
`spatial_list`	tagged list of locatoinal information from , i.e., from `FishStatsUtils::make_spatial_info`

`ObsModel_ez` an optional matrix with two columns where first column specifies the distribution for positive catch rates, and second element specifies the functional form for encounter probabilities

ObsModel_ez[e,1]=0
: Normal

ObsModel_ez[e,1]=1
: Lognormal

ObsModel_ez[e,1]=2
: Gamma

ObsModel_ez[e,1]=3
: Inverse-Gaussian

ObsModel_ez[e,1]=5
: Negative binomial

ObsModel_ez[e,1]=6
: Conway-Maxwell-Poisson (likely to be very slow)

ObsModel_ez[e,1]=7
: Poisson (more numerically stable than negative-binomial)

ObsModel_ez[e,1]=8
: Compound-Poisson-Gamma, where the expected number of individuals is the 1st-component, the expected biomass per individual is the 2nd-component, and SigmaM is the variance in positive catches (likely to be very slow)

ObsModel_ez[e,1]=9
: Binned-Poisson (for use with REEF data, where 0=0 individual; 1=1 individual; 2=2:10 individuals; 3=>10 individuals)

ObsModel_ez[e,1]=10
: Tweedie distribution, where epected biomass (lambda) is the product of 1st-component and 2nd-component, variance scalar (phi) is the 1st component, and logis-SigmaM is the power

ObsModel_ez[e,1]=11
: Zero-inflated Poisson with additional normally-distributed variation overdispersion in the log-intensity of the Poisson distribution

ObsModel_ez[e,1]=12
: Poisson distribution (not zero-inflated) with log-intensity from the 1st linear predictor, to be used in combination with the Poisson-link delta model for combining multiple data types

ObsModel_ez[e,1]=13
: Bernoilli distribution using complementary log-log (cloglog) link from the 1st linear predictor, to be used in combination with the Poisson-link delta model for combining multiple data types

ObsModel_ez[e,1]=14
: Similar to 12, but also including lognormal overdispersion

ObsModel_ez[e,2]=0
: Conventional delta-model using logit-link for encounter probability and log-link for positive catch rates

ObsModel_ez[e,2]=1
: Alternative "Poisson-link delta-model" using log-link for numbers-density and log-link for biomass per number

ObsModel_ez[e,2]=2
: Link function for Tweedie distribution, necessary for `ObsModel_ez[e,1]=8` or `ObsModel_ez[e,1]=10`

ObsModel_ez[e,2]=3
: Conventional delta-model, but fixing encounter probability=1 for any year where all samples encounter the species

ObsModel_ez[e,2]=4
: Poisson-link delta-model, but fixing encounter probability=1 for any year where all samples encounter the species and encounter probability=0 for any year where no samples encounter the species

`OverdispersionConfig`
: a vector of format c("eta1"=0, "eta2"="AR1") governing any correlated overdispersion among categories for each level of v_i, where eta1 is for encounter probability, and eta2 is for positive catch rates, where 0 is off, "AR1" is an AR1 process, and >0 is the number of elements in a factor-analysis covariance (by default, c("eta1"=0, "eta2"=0) and this turns off overdispersion)

`RhoConfig`
: vector of form c("Beta1"=0,"Beta2"=0,"Epsilon1"=0,"Epsilon2"=0) specifying whether either intercepts (Beta1 and Beta2) or spatio-temporal variation (Epsilon1 and Epsilon2) is structured among time intervals (0: each year as fixed effect; 1: each year as random following IID distribution; 2: each year as random following a random walk; 3: constant among years as fixed effect; 4: each year as random following AR1 process); If missing, assumed to be zero for each element

`VamConfig`
: Options to estimate interactions, containing three slots:

  VamConfig[0]
  : selects method for forming interaction matrix; Turn off feature using 0, or I recommend using 2 by default

  VamConfig[1]
  : indicates the rank of the interaction matrix, indicating the number of community axes that have regulated dynamics

  VamConfig[2]
  : Indicates whether interactions occur before spatio-temporal variation (`VamConfig[2]=0`) or after `VamConfig[2]=1`

`Aniso`
: whether to assume isotropy (Aniso=0) or geometric anisotropy (Aniso=1)

`PredTF_i`
: OPTIONAL, whether each observation i is included in the likelihood (PredTF_i[i]=0) or in the predictive probability (PredTF_i[i]=1)

`Xconfig_zcp`
: OPTIONAL, 3D array of settings for each dynamic density covariate, where the first dimension corresponds to 1st or 2nd linear predictors, second dimension corresponds to model category, and third dimension corresponds to each density covariate

  Xconfig_zcp[z,c,p]=0
  : `X_itp[,,p]` has no effect on linear predictor z for category c

  Xconfig_zcp[z,c,p]=1
  : `X_itp[,,p]` has a linear effect on linear predictor z for category c

  Xconfig_zcp[z,c,p]=2
  : `X_itp[,,p]` has a spatially varying, zero-centered linear effect on linear predictor z for category c

  Xconfig_zcp[z,c,p]=3
  : `X_itp[,,p]` has a spatially varying linear effect on linear predictor z for category c

`covariate_data`
: data frame of covariate values with columns `Lat`, `Lon`, and `Year`, and other columns matching names in `formula`; `Year=NA` can be used for covariates that do not change among years (e.g., depth)

`formula`
: an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted. Similar specification to `stats::lm`

`Q_ik`
: matrix of catchability covariates (e.g., measured variables affecting catch rates but not caused by variation in species density) for each observation i

`F_ct`
: matrix of fishing mortality for each category c and year t (only feasible when using a Poisson-link delta model and specifying temporal structure on intercepts, and mainly interpretable when species interactions via VamConfig)

`t_yz`
: matrix specifying combination of levels of `t_iz` to use when calculating different indices of abundance or range shifts

`CheckForErrors`
: whether to check for errors in input (NOTE: when CheckForErrors=TRUE, the function will throw an error if it detects a problem with inputs. However, failing to throw an error is no guaruntee that the inputs are all correct)

`yearbounds_zz`
: matrix with two columns, giving first and last years for defining one or more periods (rows) used to calculate changes in synchrony over time (only used if `Options['Calculate_Synchrony']=1`)

`Options`	a vector of form c('SD_site_logdensity'=FALSE,'Calculate_Range'=FALSE,'Calculate_effective_area'=FALSE,'Calculate_Cov_SE'=FALSE,'Calculate_Synchrony'=FALSE,'Calculate_proportion'=FALSE), where Calculate_Range=1 turns on calculation of center of gravity, and Calculate_effective_area=1 turns on calculation of effective area occupied
`Expansion_cz`	matrix specifying how densities are expanded when calculating annual indices, with a row for each category c and two columns. The first column specifies whether to calculate annual index for category c as the weighted-sum across density estimates, where density is weighted by area ("area-weighted expansion", `Expansion[c,1]=0`, the default) or where density is weighted by the expanded value for another category ("abundance weighted expansion" `Expansion[c1,1]=1`). The 2nd column is only used when `Expansion[c1,1]=1`, and specifies the category to use for abundance-weighted expansion, where `Expansion[c1,2]=c2` and `c2` must be lower than `c1`.
`Z_gm`	matrix specifying coordinates to use when calculating center-of-gravity and range-edge statistics. Defaults to eastings and northings for each knots or extrapolation-grid cell.
`Version`	a version number; If missing, defaults to latest version using `FishStatsUtils::get_latest_version(package="VAST")`
`...`	interface to pass deprecated inputs, included for backwards compatibility with previous versions which specified elements of `spatial_list` individually instead of as a single object

Value

Object of class `make_data`, containing inputs to function `VAST::Build_TMB_Fn()`