**Actions: data prep and exploration**

|  |  |  |  |
| --- | --- | --- | --- |
| actions | who | by when | file name |
|  |  |  |  |
| 1. data prep script including the 0 derivation | Isidora | draft uploaded in teams files because I do not have push access to github. still needs to be turned into a full function that can be sourced as is | initial\_data\_prep.r |
| 1.1. delete NAs for spatial explanatory variables; double check NAs for lat and long and delete cases where NAs in lat long and c-square |  |  |  |
| 1.2. c squares that never get fished by any metier, change those NAs with something like 'no fishing' |  |  |  |
| 1.3. 0 fishing derived from c-squares where we have env variables, these could have fishing > 0 from other metires or not, and fishing from the selected metier is 0 or NA |  |  |  |
| 2. response and expl variables exploration script | isidora |  |  |
| 3. mapping scripts of explanatory vars | serra |  |  |
| 4. Mat has some maps for fishing effort. M. can share that script |  | done already, shared in github |  |
| 5. exploration of spatial and temporal autocorrelation with acf and variograms | isidora |  |  |
| 6. plots of the relationships between response and explanatory, can build upon T's script |  |  |  |
|  |  |  |  |

**Notes**

* <https://github.com/ices-eg/wg_WGSFD/tree/master/TOR%20D>
* month vs season pros and cons:
  + seasons are a covariate that relate to environment but fleets can change their behavior inside a season this also relates to autocorrelation and how it will be modelled. Tony has used year as a time step.
  + to make train/test subsets: M. said we can follow c-squares randomly selected in time. the alternative is to randomly select observations, which might mix spatial and temporal autocorrelation which might make the estimation harder.
  + at final stage: to be done at a monthly scale?
* to make train/test subsets: M. said we can follow c-squares randomly selected in time. the alternative is to randomly select observations, which might mix spatial and temporal autocorrelation which might make the estimation harder.

decisions on data prep:

delete NAs for spatial explanatory variables; double check NAs for lat and long and delete cases where NAs in lat long and c-square

c squares that never get fished by any metier, change those NAs with something like 'no fishing'

0 fishing derived from c-squares where we have env variables, these could have fishing > 0 from other metires or not, and fishing from the selected metier is 0 or NA

**Question to ICES data center:**

* why rounded fishing effort values (rounded to hours)
* why very high fishing effort values (e.g. OT\_CRU > 200h)