Do species **revisit the same places** in successive seasons?
A **new metric** can help us find out.

Using the earth mover's distance to assess changes in species' spatial distribution

Jan Zouhar, Zuzana Musilová, Petr Musil

INTRO

To identify changes in spatial distribution across a given region (e.g., a country), it helps to have a measure of how similar the distribution was in (say) successive years. We advocate the use of the **earth mover's distance** (**EMD**) for monitoring data. Advantages of this metric include its **meaningful interpretation** and the **availability of tools** that facilitate its calculation for bird monitoring data **in R**.

~~EARTH~~ BIRD MOVER'S DISTANCE?

The **average distance an individual would have to travel** if we were to turn this year's distribution of counts across monitored sites (**Table 1**) into that from the last year **with minimum effort**.

CALCULATION

We need an **optimal flight plan**: one that **minimizes the total kilometrage** the birds need to fly to turn one spatial distribution into the other (**Figure 1, Table 2**). In operations research, this is known as the **transportation problem**. Several **polynomial-time** algorithms exist, some of them implemented **in R**. We used the **emdist** package, combined with **geosphere** to calculate site distances based on their latitude and longitude (**Figure 2**).

APPLICATION EXAMPLE

Figure 3 presents the Czech IWC data for the Great Cormorant that show a **change in spatial distribution** between 2017 and 2018; the change is reflected by a **bump in EMD values**. **Figure 4** shows a long-term time series of EMD values of six waterbird species based on Czech IWC data over the last 20 years.

LIMITATIONS

- EMD is certainly not the average distance between an individual's location in 2021 and 2022: this would require data on marked individuals. (EMD is an estimate of its lower bound, though.)
- Relatedly, **total abundance** of species **affects EMD** to a certain degree. This needs to be noted especially in interspecies comparisons.
- Incomplete and **time-varying monitoring coverage adds noise**.
- EMD is inherently **restricted to** a given **monitoring region**.
- EMD calculation may get time-consuming. In a species with over 500 occupied monitoring sites, our computational time was under 20 sec. In larger datasets, approximation by Euclidean distance may help.

EXTENSIONS

It is straightforward to **extend EMD beyond year-to-year comparisons**. Examples include (i) yearly deviations from long-term average distributions or (ii) inter-species analyses.

Table 1: An example of monitoring data for EMD calculation

Site	Latitude (*lat*)	Longitude (*lon*)	Count in 2022 (*count*)	Count in 2021 (*lag_count*)
Láska (Love)	49° 6' 35.460'' N	14° 44' 54.6792'' E	160	80
Prkenný (Wooden)	49° 6' 43.956'' N	14° 47' 10.0176'' E	0	108
Klec (Cage)	49° 5' 23.438'' N	14° 46‘ 00.4944'' E	28	0

Figure 1: Optimal flight plan for Table 1

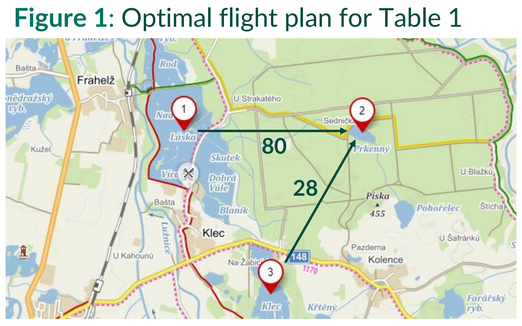


Table 2: Distances in the optimal flight plan

	1 → 2	3 → 2	Still	Total
Distance (km)	2.757	2.594	0	
Individuals	80	28	80	188
Total dist. (km)	220.56	72.63	0	293.19
EMD (avg. dist.)				**1.560**

Figure 2: R code that transforms data from Table 1 for use with **emdist::emd()**

```
calculate_emd <- function(df, max.iter = 1e5) {
   # Filter and standardize monitoring data for both years
   df <- df %>%
      filter(!is.na(count) & !is.na(lag_count)) %>%  # Keep only sites observed in both years
      filter(count > 0 | lag_count > 0) %>%  # Drop sites with 0s in both years
      mutate(  # Standardize counts into probability distributions
         count     = count     / sum(count),
         lag_count = lag_count / sum(lag_count)
      )
   A <- df %>% select(count,     lat, lon) %>% as.matrix()  # Input for emdist::emd()
   B <- df %>% select(lag_count, lat, lon) %>% as.matrix()  # Ditto
   # Calculate the EMD
   emdist::emd(A, B, max.iter=max.iter, dist=function(x, y) geosphere::distGeo(y[1:2], x[1:2]))
}
```

Figure 3: Spatial distribution and EMD for the Great Cormorant (IWC, Czechia, 2017–2019)

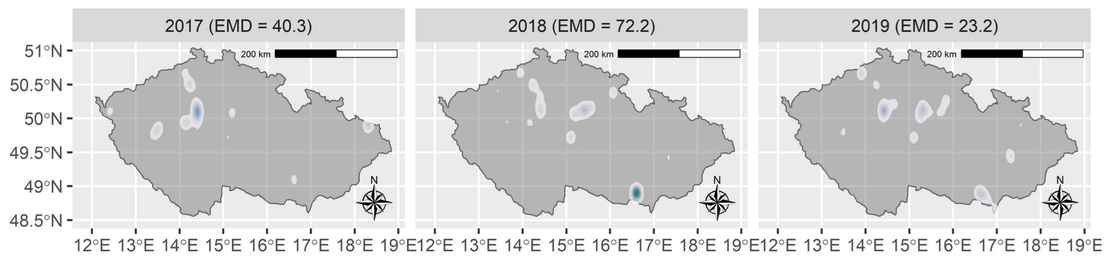
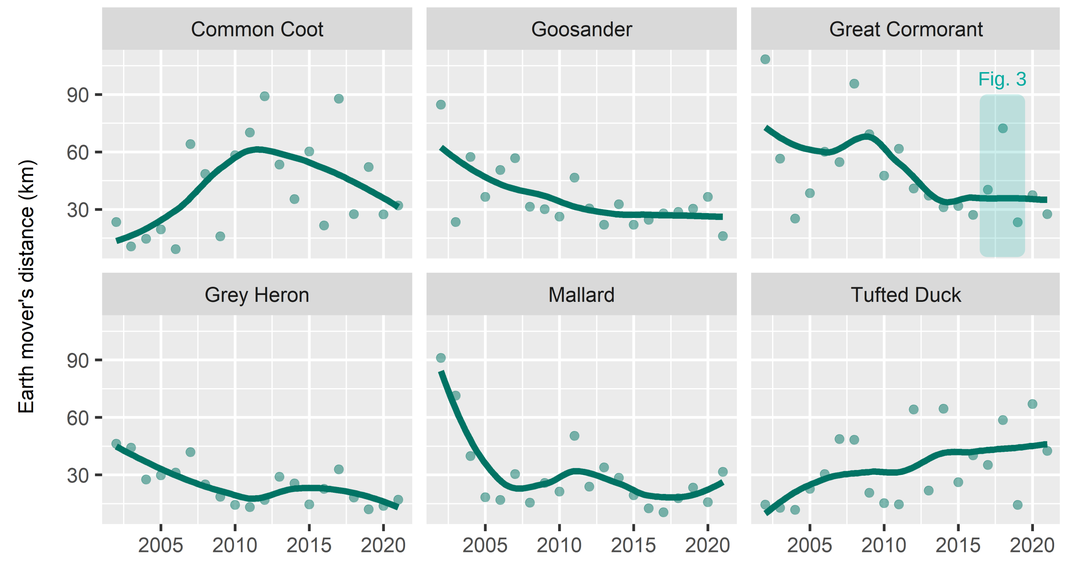


Figure 4: EMD values for selected waterbird species (IWC, Czechia, 2001–2021)



T A
Č R

Czech University of Life Sciences Prague

Bird Numbers 2022
4–8 April 2022
Lucerne, Switzerland

Take a picture to download the poster