

Advancing Ecological Informatics: A Tool for Implementing the Abundance–Preference Approach in Assemblage Analysis

Mohamed El Farouki^{1*}, Mohamed Dakki², Meryem Ennakri³, Soumia Ziti⁴

Intelligent Processing and Security of Systems-Faculty of Sciences,
University Mohammed Vth of Rabat, 10106 Rabat, Morocco^{1, 3, 4}

Scientific Institute (Zoology and Animal Ecology), University Mohammed Vth of Rabat, 10106 Rabat, Morocco²

Abstract—Although the Abundance-Preference Diagram (APD) was introduced four decades ago as an entropy-based framework for classifying species along habitat-preference gradients, its uptake by the ecological community has remained marginal. The cause is not theoretical but operational: no existing software (e.g., vegan, ade4, CANOCO, PC-ORD, BIOMAPPER) implements the Degree of Preference index (DP) or the APD, which forces ecologists either to rebuild the algorithm from scratch or to abandon the method altogether. This barrier mirrors a broader concern in ecoinformatics, where the gap between published methods and accessible implementations continues to slow the adoption of quantitative tools by field ecologists. This study presents EcoAPD (<https://apd.biodiversity.ma>), the first open web-based platform dedicated to the APD method. The application takes a standard species-by-habitat abundance matrix and returns the full diagnostic suite (DP values, exclusive species curve, hyperbola-based preference classes, and the APD plot itself), without any programming or local installation. Three design choices set it apart from the descendants of CANOCO and ade4: 1) a dual-computation architecture that generates the normalized and non-normalized APD versions simultaneously at project creation, removing the need for users to commit to a method a priori; 2) a hybrid file/relational-database persistence layer with hash-based recompute avoidance, enabling visualizations to be regenerated without rerunning the full pipeline; and 3) a validation engine tailored to the irregularities of ecological matrices. The platform is evaluated on two real datasets and a synthetic size sweep. Vectorizing the preference-index core yields a speedup rising from 3.5× at 10 species to 85.6× at 1000 species, while full-pipeline latency is ~510 ms on a 93-species × 37-habitats matrix and ~2.3 s on a 114-species × 171-habitats waterbird monitoring matrix, the residual cost being dominated by habitat count rather than species count. A threshold-sensitivity ablation finds the classification stable at the 0.3 and 0.75 boundaries and sensitive at 0.1 and 0.5. EcoAPD makes a four-decade-old method operationally available, with potential to revisit historical assemblage studies that relied on abundance data alone.

Keywords—Ecological informatics; web application; algorithm; assemblage ecology; species preference; abundance-preference diagram

I. INTRODUCTION

The preference of a species in a given environment is a key concept in both ecology and biodiversity conservation [1],[2],[3]. It is generally compiled from the abundance

distribution of a species in different habitats [4]. The degree of preference (DP), as defined by [5], is an index that determines whether a species is a generalist or a specialist, in terms of its habitat preference, without necessarily indicating which habitats it prefers. This index shows high efficiency in analyzing assemblage organization through the APD method, which combines DP and species abundances [6],[7].

Although first published four decades ago, this method remained underutilized due to the lack of accessible tools for ecologists, who are mostly not familiar with statistical tools, a challenge that persists due to the frequent mismatch between ecological data and traditional statistical knowledge and tools [8],[9]. The recent refinements of the method [6] and the current abundance of efficient informatics tools have now made its widespread implementation feasible.

Our research consists of translating this method into a practical tool: a web application that executes all necessary calculations and plots the diagnostics. This application utilizes a standard data matrix of species abundances across a sample of habitats and covers both the original and normalized versions of the APD [7].

The development of this application was preceded by a comparative analysis of available algorithms for processing data in the form of contingency tables, and of the software platforms that analyze taxonomic and ecological data. That analysis, reported in Section II, establishes that none of the platforms in current use can implement the DP/APD tools. The platforms that come closest are separated from their intended users by either a programming requirement or a commercial licence.

To help ecologists implement DP/APD tools, we have previously developed an algorithm [7], and in parallel we launched a first version of the EcoAPD web application, which has been significantly improved since then.

The contributions of this work are the following:

- EcoAPD is, to the authors' knowledge, the first open, web-based implementation of the Degree of Preference index and the Abundance–Preference Diagram, a method that no existing ecological software package implements (Section II, Table I).
- A dual-computation architecture derives both the normalized and the non-normalized APD at project

*Corresponding author.

creation, so that the choice between the two versions becomes a post-hoc display decision rather than an a priori methodological commitment (Section V-A).

- A hybrid file/relational persistence layer with hash-based recompute avoidance stores all computed intermediates, so that visualizations are regenerated without rerunning the pipeline, and re-uploading unchanged data does not re-trigger computation [Section V-A(2c)].
- The preference-index computation, originally a per-species loop, has been rewritten as a single vectorized NumPy operation, measured at a speedup rising to $85.6\times$ at 1000 species (Section VI-B, Table IV).
- An empirical evaluation reports the scalability of the full pipeline on two real datasets, together with a two-part ablation quantifying the contribution of the vectorized core and of the four classification thresholds individually (Section VI).

II. RELATED WORK

The platforms available for the analysis of taxonomic and ecological data organized as contingency tables are examined here along three axes: the ecological question they address, their accessibility profile, and whether they expose any entropy-based habitat-preference index.

A. Methods for Habitat Preference and Assemblage Structure

BIOMAPPER [10],[11] implements ecological-niche factor analysis (ENFA), which characterizes the niche of a single species by contrasting the environmental conditions at occupied sites with the global distribution of those conditions.

MaxEnt [12] addresses a related problem, species distribution modelling from presence-only records, by fitting a maximum-entropy distribution over environmental covariates. Both are concerned with where a species occurs as a function of environmental predictors; neither takes a species-by-habitat abundance matrix as its primary input, and neither produces a

per-species preference index derived from the entropy of the abundance distribution itself.

The ordination family addresses assemblage structure rather than single-species niches. CANOCO [17] and PC-ORD [16] provide constrained and unconstrained ordination behind graphical interfaces, while vegan [13] and ade4 [15] provide the same method family, and a considerably wider one, as R packages. These methods summarize the joint variation of species and sites on a small number of latent axes, and a species' position on those axes reflects its association with the sites that load on them. This is informative, but it is not a preference index: the axis scores are defined relative to the ordination solution for the table as a whole, they change when species or sites are added or removed, and they do not map onto an interpretable scale bounded between generalist and specialist.

iNEXT [18] and the Hill-number framework [22] quantify diversity at the level of the assemblage rather than of the individual species, and are concerned with sampling completeness rather than with habitat preference. ProCURVE [14] computes habitat preferences, but for aquatic organisms specifically and from a different formulation. GBIF [19] is an occurrence-data infrastructure rather than an analysis platform.

The DP index [5] differs from all of these in that it is computed independently for each species, from the entropy of that species' own abundance profile across habitats, normalized by $\log_2(J)$. It is bounded, it is invariant to the composition of the rest of the table, and it admits a direct reading: a value near 0 for a species distributed evenly across all habitats, approaching 1 for a species confined to a single one. None of the platforms surveyed computes this quantity, and none produces the APD. Table I summarizes the comparison.

The negative entries in the DP and APD columns of Table I are not a shortcoming of these packages, each of which is well established within its own method family; they are precisely the gap that motivates the present work.

TABLE I. COMPARISON OF ECOLOGICAL ANALYSIS PLATFORMS WITH RESPECT TO THE DP INDEX AND THE APD

Platform	Primary method family	DP index	APD plot	Programming required	Installation	Licence
BIOMAPPER [10],[11]	ENFA / niche factor analysis	No	No	No	Desktop	Free
MaxEnt [12]	Maximum-entropy SDM	No	No	No	Desktop (Java)	Free
vegan [13]	Ordination, diversity	No	No	Yes (R)	R package	GPL
ProCURVE [14]	Aquatic habitat preference	No	No	No	Desktop	Free
ade4 [15]	Multivariate / duality diagram	No	No	Yes (R)	R package	GPL
PC-ORD [16]	Multivariate community analysis	No	No	No	Desktop	Commercial
CANOCO [17]	Constrained ordination	No	No	No	Desktop	Commercial
iNEXT [18]	Diversity interpolation/extrapolation	No	No	Yes (R) or web	R package or web	Free
EcoAPD (this work)	Entropy-based preference classification	Yes	Yes	No	None (web)	Free

B. Accessibility as a Methodological Constraint

The second axis is orthogonal to method, and concerns who is able to run it. vegan and ade4 are free, actively maintained and

computationally capable, but they presuppose fluency in R: a user must be able to import a matrix, call the appropriate function with the appropriate arguments, and interpret an object returned at the console. CANOCO and PC-ORD remove that

requirement through graphical interfaces, but are commercially licensed and locally installed, which places them beyond the reach of under-resourced institutions. BIOMAPPER is free and graphical, but restricted to ENFA and no longer actively developed.

This matters because the population that would use the APD method consists largely of field ecologists whose training is taxonomic and ecological rather than computational. Quantitative training among early-career ecologists is known to

be limited [8], and the difficulty of translating published ecological methods into runnable code is a recognized bottleneck in ecoinformatics [9],[20]. Accessibility is therefore not a convenience feature but a determinant of whether a method is used at all, which is the most plausible explanation for the APD remaining dormant for four decades despite being theoretically available throughout.

TABLE II. LIMITATIONS OF EXISTING PLATFORMS AND THE CORRESPONDING ECOAPD DESIGN DECISIONS

Limitation identified in existing tools	Consequence for the ecologist	EcoAPD design decision
No DP/APD implementation in any surveyed platform	Method unusable without bespoke coding	Direct implementation of (1)–(4)
Programming fluency presupposed (vegan, ade4)	Excludes non-computational ecologists	Browser interface, no installation
Commercial licence (PC-ORD, CANOCO)	Excludes under-resourced institutions	Free and openly accessible
Method version must be chosen before analysis	Commitment made without sight of the results	Dual computation at project creation
Re-analysis requires full recomputation.	Iterative exploration is costly	Hybrid file/relational persistence with hash-based recompute avoidance
Input assumed clean and regular.	Real survey matrices fail, sometimes silently	Dedicated validation engine

C. Design Requirements Derived from Existing Tools

Read together, the two axes above yield a set of requirements that shaped the architecture of EcoAPD. Table II maps each limitation identified above onto the design decision it motivated.

Two of these decisions warrant emphasis, since they are what distinguish EcoAPD from a straightforward web wrapper around the published algorithm. The first is dual computation: because both APD versions are derived at project creation, the user is never required to choose between them before seeing either, which removes the analytical commitment overhead borne by users unfamiliar with the method. The second is hash-based recompute avoidance in the persistence layer: a hash of the abundance matrix together with its column mapping identifies whether an uploaded dataset is materially unchanged, so that reuploading the same data does not re-trigger the pipeline.

III. REMINDER: APPLICATION OBJECT (DP AND APD)

The APD method consists of analyzing flora or fauna assemblages using a diagram in which the species are positioned according to both their abundance, as a common descriptor in ecology, and their degree of preference (DP). The latter is measured for each species using an adapted form of the Shannon entropy formula [21], calculated on the series of its abundances in a set of habitats. The APD contains a significant reference curve, the exclusive species curve (ESC), and is subdivided into six preference zones (or classes) by five half-hyperbolas of the same form, defined by the equation $Y = (ax+b)/(cx+d)$ Fig. 1. Four of these curves have their vertices on ESC points of significant abscissa (DP), namely 0.1, 0.3, 0.5 and 0.75; these values were defined on the basis of experimental datasets. The fifth curve, which delimits the Accidental class in the non-normalized version, is not defined by a DP target but is derived from the habitat count and a theoretical abundance scale (Section V-E).

In practice, the data on which the main calculations are based are presented in a table of I lines (species) and J columns

(habitats), which is commonly used in assemblage studies. The APD method consists of calculating the DP for each species, using the following formula. Then, the APD is simply plotted for each assemblage (= habitat).

$$D_{ij} = 1 - \frac{S_i}{\log_2(J)} \quad (1)$$

$$S_i = - \sum \frac{n_{ij} + 1}{\sum (n_{ij} + 1)} \log_2 \frac{n_{ij} + 1}{\sum (n_{ij} + 1)} \quad (2)$$

n_{ij} : abundance of species i in habitat j; $\log_2(J)$: maximum value of S_i (same abundance value (n_{ij}) in all habitats; D_{ij} : degree of preference (DP).

For an exclusive species (which exists only in one habitat), Formula (1) becomes as follows:

$$S_i = \frac{n_{ij} + 1}{n_{ij} + J} \left(\frac{n_{ij} + 1}{n_{ij} + J} \right) + (J - 1) * \frac{1}{n_{ij} + J} \left(\frac{1}{n_{ij} + J} \right) \quad (3)$$

In the normalized version of the APD [7], mainly used when abundances are low, the Y-axis (abundances) is transformed using Formula (4); then, all the other components of the APD, including the species positions, are recalculated. In this version, we chose another hyperbola formula, $Y = a^2/2x$, while we used only five hyperbolas, the accidental species being interpreted as foreign to the assemblage.

$$A_{ii} = \frac{\log_2(a_{ij} + 1)}{\log_2(\max(a_{ij}) + 1)} \quad (4)$$

a_{ij} = abundance of the species i in the assemblage j;

$\max(a_{ij})$ = the maximum abundance recorded across the entire species $\times$ habitat matrix, i.e. a single global value and not a per-species maximum.

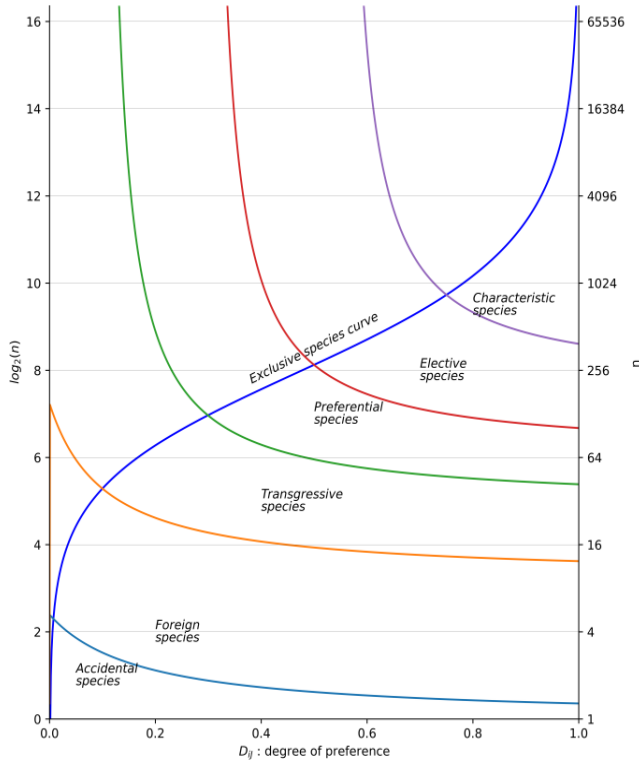


Fig. 1. APD model for assemblage analyses [5].

IV. METHODS

A. Inputs and Outputs of the Application

1) *Inputs: data and their format*: The basic data in each study case (named here 'project') are stored in a single table of I rows (species) and J columns (habitats) as indicated below (Table III). This matrix, commonly used in the ecology of assemblages [15],[16],[22], contains two types of identifiers of both rows (species name and code) and columns (habitat name and code), which respectively fill the two first columns and the two first rows, the abundances (data) are readable starting from the cell 'row 3, column 3'.

Data format, which is verified by the system, should respect the following conditions:

- input table contains a minimum of 10 habitats ($J > 9$) and at least one species ($I \geq 1$);
- data are numerical (natural or decimal numbers) and positive; empty cells are filled with '0', and empty rows or columns are deleted; for decimal numbers, only a dot '.' is accepted as separator of decimals, while separators of thousands are not accepted.

In addition, the habitat and species names should be easily recognized by the project holder, while their codes are shorter, as strings or numbers, and very useful for displaying species in the APD, avoiding overlapping of long names.

TABLE III. GENERAL FORMAT OF THE BASIC DATA MATRIX FOR IMPLEMENTING THE APD METHOD

		Habitat 1	...	Habitat j	...	Habitat J
	Codes	H1	...	Hj	...	HJ
Species 1	S1	n1.1	...	n1.2	...	n1.J
....		...				
Species i	Si	ni.1	...	n2.2	...	n2.J
....		...				
Species I	SI	nI.1	...	nI.2	...	nI.J

Species name: names that will be displayed in the APD; **Codes**: short but meaningful identifiers that will be displayed; **I**: number of species, non-limited number; **J**: number of habitats, preferably greater than 10; **nij**: abundance of the species i in the habitat j, as a positive real number.

2) *Outputs*: The outputs are defined in consideration of the major users' needs from the EcoAPD application, which consists of plotting, for each habitat, the assemblage APD in its two versions (original and/or normalized), with the species list of each preference class. However, the system also displays the data table, including the DP values for each species Fig. 2, mainly for verification and downloading purposes.

In addition, users can generate the histogram of the species DP values, as well as the abundance profile of a selected species across all habitats. These diagrams help to understand the position of the species in the different assemblage APD.

EcoAPD

Abundance-Preference Diagram Analysis

M.El Farouki

MF

Project: Waterbird Assemblage 2018-2022

Data matrix: 113 Species x 168 Habitats

← Back to Projects

Select visualization...

Species	↑	Codes	1003	1005	1007	1010	1020	1025	1027	1030	1031	1050	1060	1065	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	1426	1427	1428	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439	1440	1441	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455	1456	1457	1458	1459	1460	1461	1462	1463	1464	1465	1466	1467	1468	1469	1470	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500	1501	1502	1503	1504	1505	1506	1507	1508	1509	1510	1511	1512	1513	1514	1515	1516	1517	1518	1519	1520	1521	1522	1523	1524	1525	1526	1527	1528	1529	1530	1531	1532	1533	1534	1535	1536	1537	1538	1539	1540	1541	1542	1543	1544	1545	1546	1547	1548	1549	1550	1551	1552	1553	1554	1555	1556	1557	1558	1559	1560	1561	1562	1563	1564	1565	1566	1567	1568	1569	1570	1571	1572	1573	1574	1575	1576	1577	1578	1579	1580	1581	1582	1583	1584	1585	1586	1587	1588	1589	1590	1591	1592	1593	1594	1595	1596	1597	1598	1599	1600	1601	1602	1603	1604	1605	1606	1607	1608	1609	1610	1611	1612	1613	1614	1615	1616	1617	1618	1619	1620	1621	1622	1623	1624	1625	1626	1627	1628	1629	1630	1631	1632	1633	1634	1635	1636	1637	1638	1639	1640	1641	1642	1643	1644	1645	1646	1647	1648	1649	1650	1651	1652	1653	1654	1655	1656	1657	1658	1659	1660	1661	1662	1663	1664	1665	1666	1667	1668	1669	1670	1671	1672	1673	1674	1675	1676	1677	1678	1679	1680	1681	1682	1683	1684	1685	1686	1687	1688	1689	1690	1691	1692	1693	1694	1695	1696	1697	1698	1699	1700	1701	1702	1703	1704	1705	1706	1707	1708	1709	1710	1711	1712	1713	1714	1715	1716	1717	1718	1719	1720	1721	1722	1723	1724	1725	1726	1727	1728	1729	1730	1731	1732	1733	1734	1735	1736	1737	1738	1739	1740	1741	1742	1743	1744	1745	1746	1747	1748	1749	1750	1751	1752	1753	1754	1755	1756	1757	1758	1759	1760	1761	1762	1763	1764	1765	1766	1767	1768	1769	1770	1771	1772	1773	1774	1775	1776	1777	1778	1779	1780	1781	1782	1783	1784	1785	1786	1787	1788	1789	1790	1791	1792	1793	1794	1795	1796	1797	1798	1799	1800	1801	1802	1803	1804	1805	1806	1807	1808	1809	1810	1811	1812	1813	1814	1815	1816	1817	1818	1819	1820	1821	1822	1823	1824	1825	1826	1827	1828	1829	1830	1831	1832	1833	1834	1835	1836	1837	1838	1839	1840	1841	1842	1843	1844	1845	1846	1847	1848	1849	1850	1851	1852	1853	1854	1855	1856	1857	1858	1859	1860	1861	1862	1863	1864	1865	1866	1867	1868	1869	1870	1871	1872	1873	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	228
---------	---	-------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----

B. Test Data Used

The application has been tested, in terms of functionalities and validity of the outputs, by some researchers that we invited to test the EcoAPD application. However, for experimenting with different development steps, we used two different real datasets, provided here as supplementary materials, named SM1 and SM2:

- SM1: Benthic invertebrates of the Sebou river;
- SM2: International winter counts of waterbirds in Morocco (average of 2018-2022 counts);

These data are borrowed from different ecological studies that provide data adapted to APD use, in both original and normalized forms.

C. Used Technologies

This application is based on the algorithm already published by [6] and normalized by [7].

The application was developed using Python [23] for server-side logic and calculations, combined with standard web technologies HTML [24], CSS [25], and JavaScript [26] for the user interface.

The main computational components include:

- Mathematical libraries: NumPy [27] and Pandas [28],[29] handle matrix operations and data manipulation, enabling efficient processing of species-habitat abundance matrices.
- Data storage: Project data, user accounts, and computed results are stored in a MySQL relational database [30], ensuring data integrity and enabling rapid retrieval of data and previous analyses.
- File handling: Uploaded datasets (Excel files) are stored locally, each assigned a unique identifier generated from the project ID and timestamp to ensure traceability and avoid naming conflicts.
- Visualizations: Matplotlib [31] and Apache ECharts [32] generate static and interactive graphs, respectively, enabling efficient processing and visualization of species-habitat abundance matrices.
- Export capabilities: The system generates publication-quality PDF reports containing APD diagrams, data tables, and project metadata.

Together, these technologies enable a streamlined workflow for ecological data analysis, ensuring both computational efficiency and a seamless end-user experience.

V. SYSTEM DESIGN AND IMPLEMENTATION

A. System Architecture

1) *Preamble*: A first version of this system was created to implement the original version of the APD tool [5]; we further submitted it for major modifications in order to implement the normalized version of this tool [6]. In addition to multiple tests of the web application done by our staff, some potential users

were asked to write essays about this system, using their own data, and their remarks were highly beneficial for improving it.

The application implements a robust database architecture, storing abundance matrices and all computed results, including DP values, ESC, hyperbola set, and species points. This persistence layer enables rapid generation and visualization without re-computing the entire pipeline, particularly beneficial for large datasets or iterative analysis workflows.

A distinctive feature of the application is its automatic dual-dataset generation: when a project is created, the system automatically processes both normalized and non-normalized APD datasets and stores them in a database Fig. 3. This design avoids initializing a method before starting analyses.

2) *Architectural components*: The application is built with technologies that allow optimizing the calculation speed, while offering a fluid user experience Fig. 4.

a) *User interface (application layer)*: The web interface is responsive, enabling dynamic interaction with the application through

- Authentication portal: Secure login and registration with email verification and reCAPTCHA protection [33];
- Project dashboard: Intuitive management of analysis projects with search and metadata editing;
- Data configuration wizard: Excel upload with preview and flexible mapping of species/habitat identifiers;

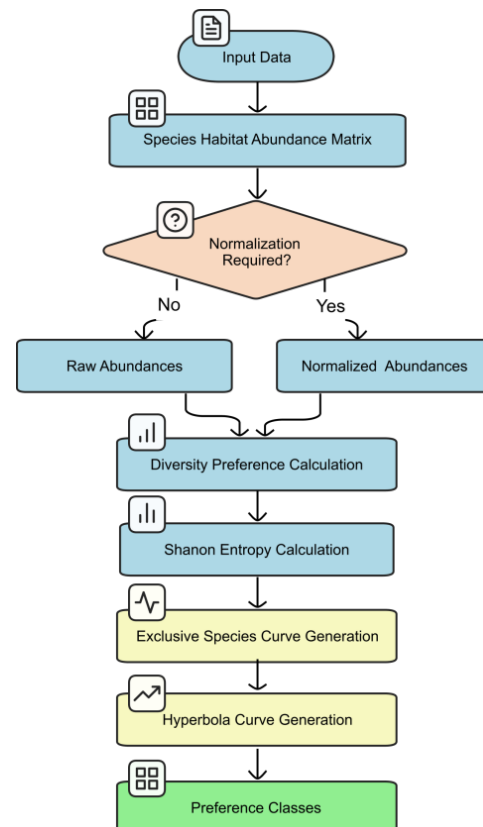


Fig. 3. Automatic generation and storage of normalized and non-normalized APD datasets during project creation.

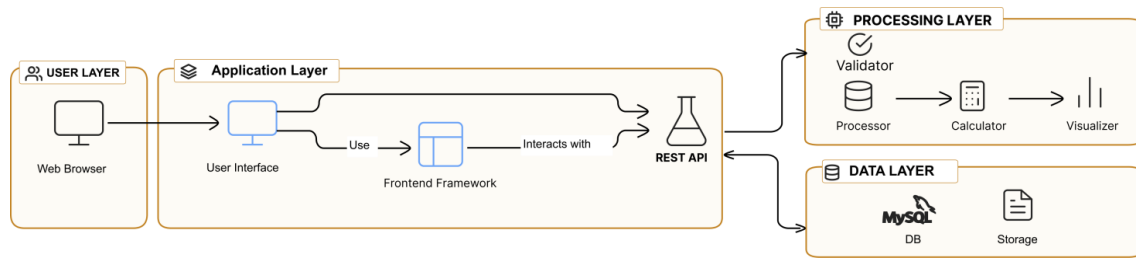


Fig. 4. System architecture diagram showing the main components and their interactions.

- interactive APD viewer: Dynamic diagrams supporting zoom, pan, and species labeling with real-time parameter adjustments;
- exploration tools: Sortable tables for abundances and DP values, species abundance profiles, and DP histograms;
- multi-format export: Download options for PNG images, and publication-ready PDF and Excel reports.

b) Server-side processing (processing layer): The server component handles requests from the user interface, processes data, and manages database interactions. It provides the following main functionalities:

- data validation engine: verifies uploaded files against format requirements, cleans invalid entries, and prepares matrices for analysis;
- DP computation module: calculates degree of preference values and generates both normalized and non-normalized datasets automatically;
- curve generation: computes ESCs and hyperbola sets for assemblage classification;
- security layer: enforces authentication, session management, and input sanitization across all operations.

All operations include appropriate security checks and error handling to ensure data integrity.

c) Data storage (data layer): The initial system, announced by [6], used pickle-based file storage to persist computed DP objects, enabling rapid visualization without recomputation; this approach provides fast state restoration but lacks query flexibility and data integrity. In the actualized version of EcoAPD, we migrated to a relational database, where Excel files are stored in the dedicated project files, while all computed data and metadata are stored in the MySQL database Fig. 5. This system, combining file-based and database storage, provides flexibility and performance, data integrity, and query capabilities. In addition, this architecture enables selective data retrieval, supports concurrent access, maintains referential integrity through foreign keys, and facilitates schema evolution Fig. 6.

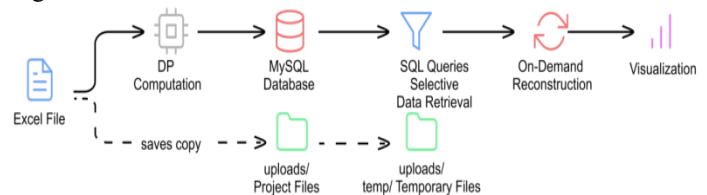


Fig. 5. Data persistence and retrieval workflow.

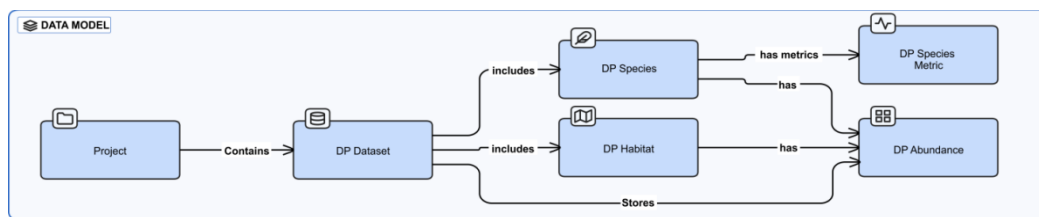


Fig. 6. Database structure showing relationships between user, project, and computed data.

d) Mathematical processing: The application uses optimized mathematical libraries that enable rapid processing of large ecological datasets. The main computations include: 1) *abundance data*: species-habitat tables are stored as NumPy arrays for fast, bulk operations; 2) *DP calculation*: Shannon entropy and related math run efficiently on entire arrays; 3) *normalization*: log2 scaling and value adjustment are done in parallel for all data; 4) *diagram data*: APD curves and coordinates are generated in batches.

B. User Workflow

The system provides a structured workflow for ecological data analysis Fig. 7, which shows the key interaction sequences

during a typical analysis session. This workflow is designed to minimize technical barriers to users unfamiliar with computational complexity, allowing them to focus on result interpretation rather than on data computation.

C. Workflow Steps: Data Processing Pipeline

The pipeline consists of several interconnected stages that transform raw biodiversity data into interpretable visualizations and classification results, which can be edited for publication Fig. 8.

The workflow can be subdivided into three main phases: 1) data preview, validation and upload, 2) selection and configuration of outputs, and 3) generating outputs.

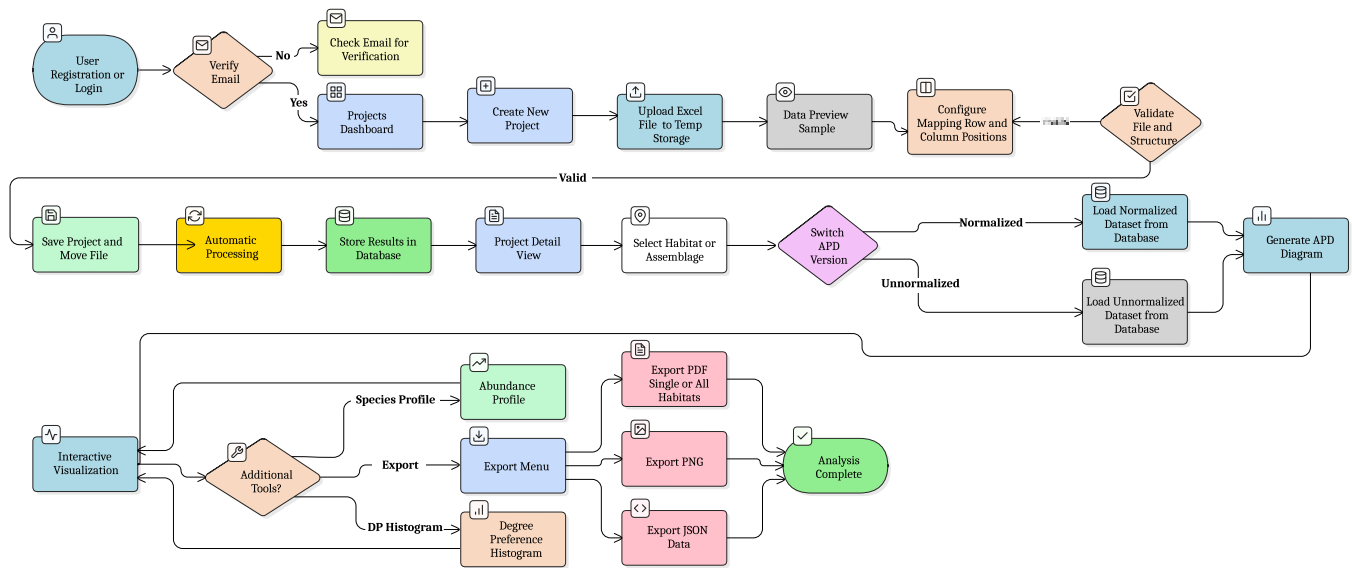


Fig. 7. User workflow diagram: key interaction sequences during a typical analysis session.

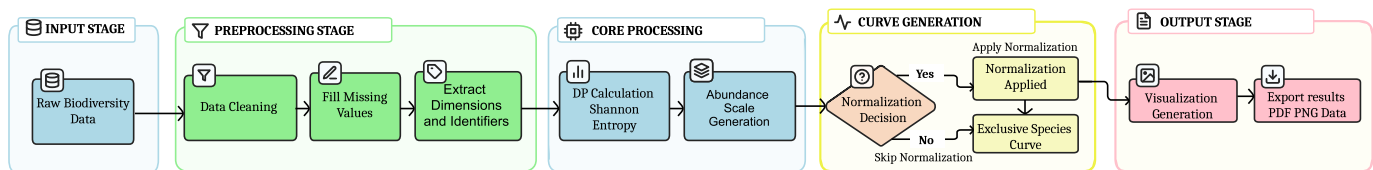


Fig. 8. Data processing pipeline architecture showing the transformation of raw biodiversity data into interpretable visualizations.

1) *Data preview, validation and upload:* The use of the APD method starts with creating a project, based on a 'species-habitat' matrix of abundances, which is provided in an Excel file. The system architecture is conceived to meet the specific requirements of computed data, knowing that ecological datasets have different sizes and complexity depending on studies. Users are previously prompted to configure their data; however, the dataset is submitted to a validation sequence that is implemented before creating the project. The first sequence consists of displaying the first '5 rows x 5 columns' of the project file Fig. 9, which allows a visual verification that names and codes of both rows and columns are provided.

2) *Selecting outputs:* Once the data is validated and uploaded, the user is requested to select different analysis types and options that can be edited:

- the most requested output is the APD, for a selected assemblage; it is edited in a desired version (normalized vs. non-normalized), where species are displayed using a selected identifier (name or code), and the reference curves (hyperbolas and ESC) are displayed or hidden Fig. 10;
- DP profile for all species Fig. 11;
- abundance profile of a selected species in all habitats Fig. 12.

3) *Generating outputs:* The requested APD is displayed and, when needed, it can be saved in image or PDF format. The

other graphical outputs related to species (abundance profiles and DP diagram) can be displayed and saved in these same formats, while the specific composition of each preference class of the APD can be edited in Excel format. All these outputs have a printable quality, directly used in a publication.

D. Workflow Implementation

The main processing flow for successfully uploading a file proceeds as follows Fig. 13.

- Once uploaded, the data file is transmitted to the API, which submits it to a file validator; this latter checks both extension (.xlsx/.xls), size (< 100 Mb), file structure, and data format.
- The file is saved in the storage directory, with a unique filename. The API then submits the data to a data processor that checks its required structure and type, deleting non-numeric values and empty rows and columns, filling empty cells with a '0' value, and extracting species and habitat labels. The matrix dimensions are then extracted, as well as row and column identifiers, while columns are converted to strings.
- Once the project is created, the system processes the data simultaneously to compute DP values, the ESC and hyperbolas for both normalized and non-normalized APD methods;
- All project metadata and computed results are stored in the database: species and habitat metadata, abundance matrices (stored in long format), DP values for each

species, ESC and their intersection points with hyperbolas, and project record creation and storage of file path and data mapping.

- At the end of this process, the API returns a success response to the web interface, which displays the data table as a confirmation of the project creation.

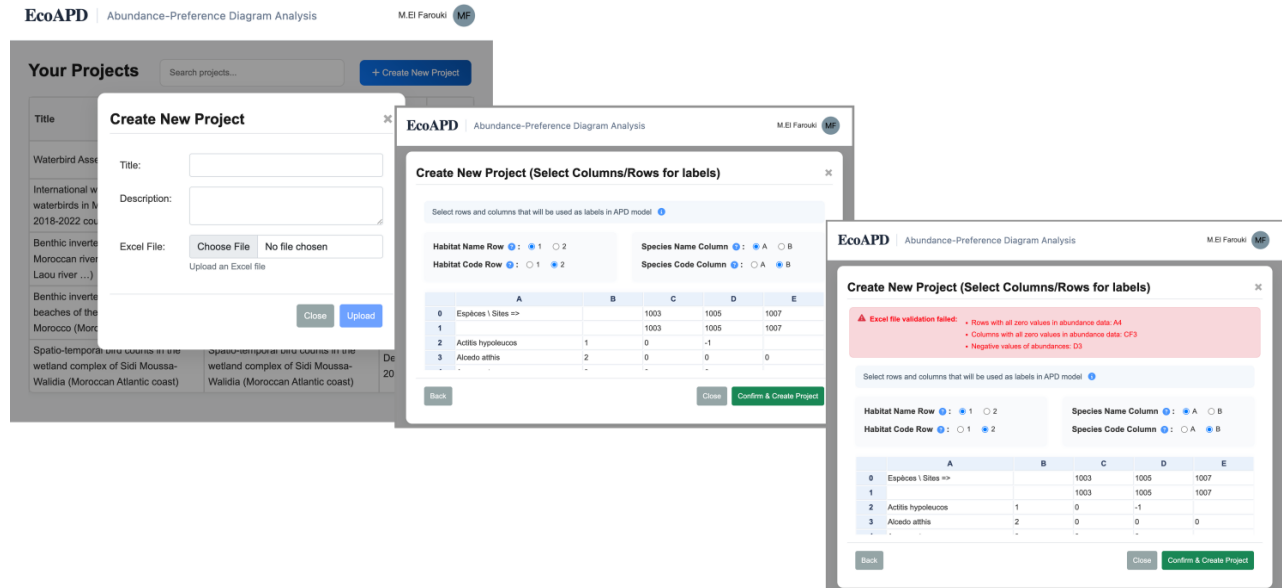


Fig. 9. Upload and validation interfaces.

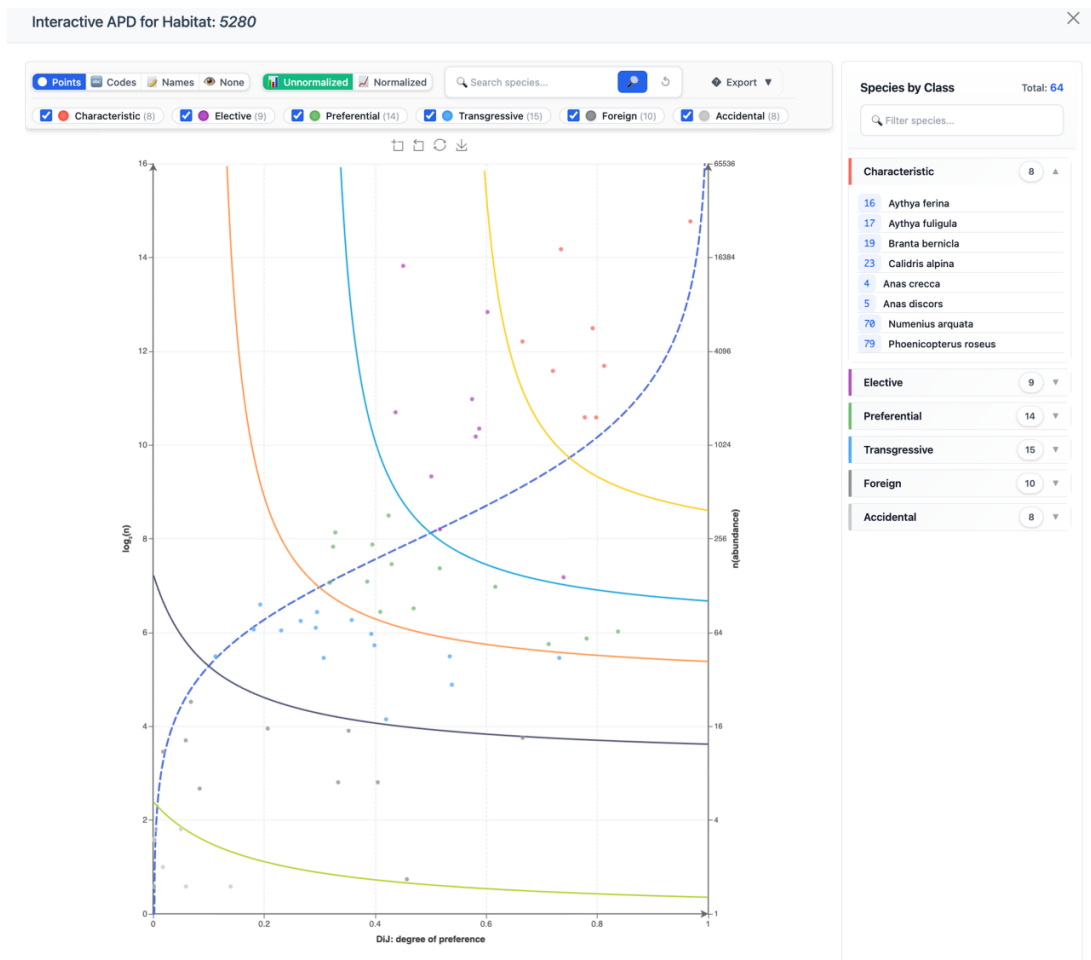


Fig. 10. APD of an assemblage/habitat: Interactive visualization.

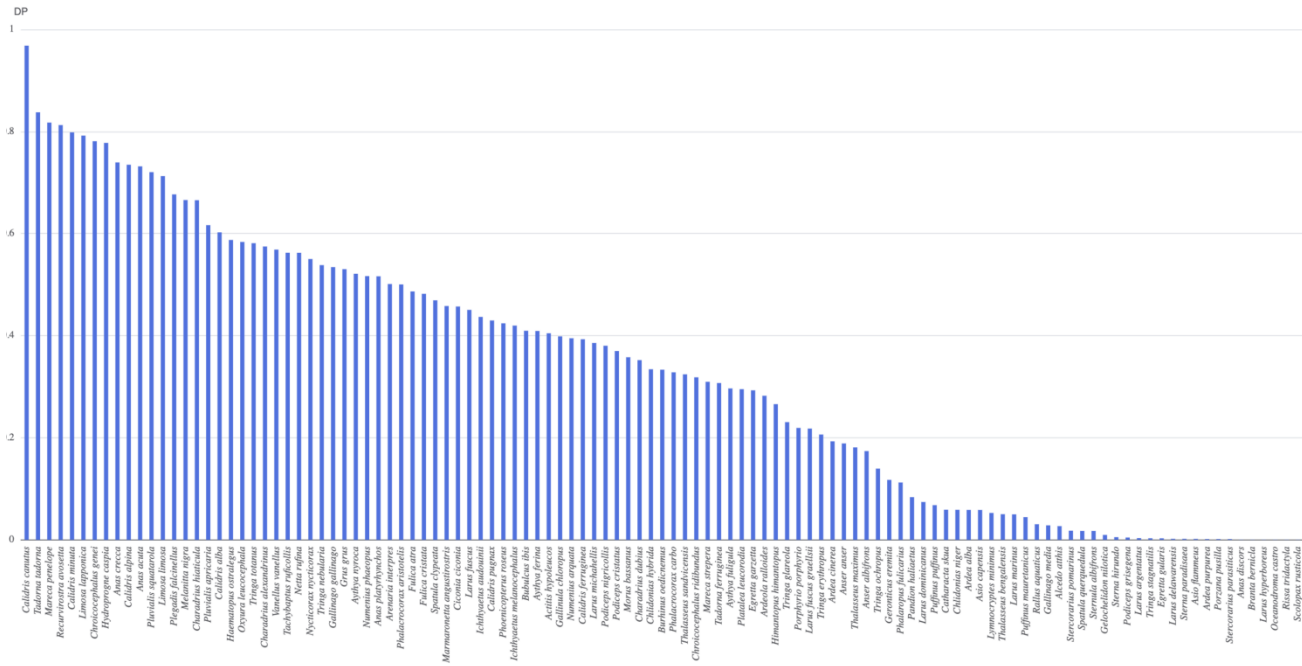


Fig. 11. DP profile for all species.

E. Generation of the Classification Curves (ESC and Hyperbolas)

Theoretical curves (ESC and hyperbolas) for species classification are generated from a computed vector. For the ESC, matrix operations start with $Y = \text{esp_array} + 1$.

The entropy and the DP are then obtained as:

$$Si = -\sum((Y/\sum(Y)) \times \log_2(Y/\sum(Y))), \text{ and the DP as } Di = 1 + Si/\log_2(S)$$

The hyperbola curves that delineate the zones corresponding to the species preference classes are generated from their intersections with the ESC, which are predefined by their abscissa (DP value). Four of these curves have vertices at fixed DP targets — 0.1, 0.3, 0.5 and 0.75 — each located by absolute-difference minimization within an asymmetric tolerance band around its target (the 0.1 vertex, for instance, is searched within [0.095, 0.15] and the 0.75 vertex within [0.70, 0.80]) rather than under a single global tolerance. The non-normalized version generates one further curve, the baseline of the Accidental class, which is not defined by a DP target but derived from the habitat count and a theoretical abundance scale; the five curves together delimit the six preference zones. In the normalized version, this baseline curve is omitted, the accidental species being interpreted as foreign to the assemblage, which leaves four curves. In both versions, a reference hyperbola is generated and duplicated by simple translation to obtain the remaining curves.

VI. EVALUATION

This section reports the computational environment and benchmarking protocol (VI.A), an ablation of the vectorized mathematical core (VI.B), a sensitivity analysis of the classification thresholds (VI.C), the scalability of the full pipeline (VI.D), and the limits of what has been evaluated (VI.E). EcoAPD is a deterministic analytical pipeline rather than

a stochastic simulation or a trained model: it has no sampling procedure, no random seed, and no free hyperparameters, so that, given a matrix, the DP values, the ESC and the hyperbola set are uniquely determined by (1)–(4). The quantities reported below are therefore timings and classification outcomes, not estimates carrying sampling error.

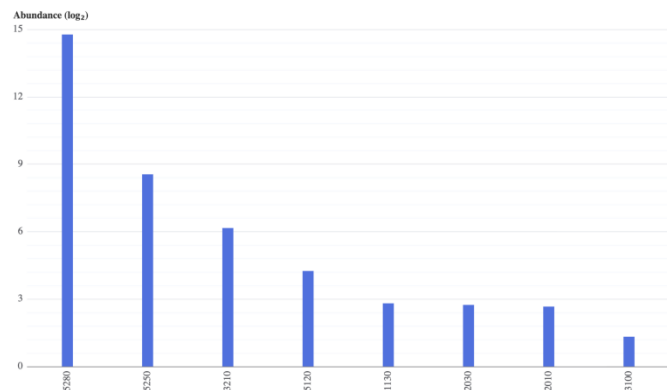


Fig. 12. Abundance profile of a selected species in all habitats.

A. Experimental Setup

Benchmarking hardware: Apple M1 Pro, 8 physical and 8 logical cores, 16 GB RAM, running macOS 26.5.2. This is the development and benchmarking machine on which the reported measurements were produced, and not necessarily the production deployment server.

Software: Python 3.9.10, NumPy 2.0.2, pandas 2.2.3, Matplotlib 3.9.2, openpyxl 3.1.5, Flask 3.0.3, Flask-SQLAlchemy 3.1.1, SQLAlchemy 2.0.36 and PyMySQL 1.1.1, all version-pinned. The database is MySQL 8.0 [30].

Timing protocol: each measurement comprises one discarded warm-up execution followed by five timed executions using `time.perf_counter()`, reported as the mean. All timings are server-side, measured by direct instrumentation of the processing layer, so that network latency and browser rendering are excluded.

Fixed parameters: the four classification-threshold targets are $DP = 0.1, 0.3, 0.5$ and 0.75 , in both the normalized and the non-normalized variants, each located within an asymmetric tolerance band around its target rather than under a single global

tolerance (Section V-E). The ESC is evaluated on $[J / 0.0015]$ points, which scales with the habitat count J rather than being fixed: 24,667 points for SM1 (37 habitats), and 114,000 for SM2 (171 habitats).

Datasets: the two real datasets SM1 and SM2 described in Section IV-B, and a synthetic series of uniform-random non-negative integer abundances, seeded at 10, 50, 100, 500, and 1000 species crossed with 5, 10, and 20 habitats, used for the size sweep of Section VI-B.

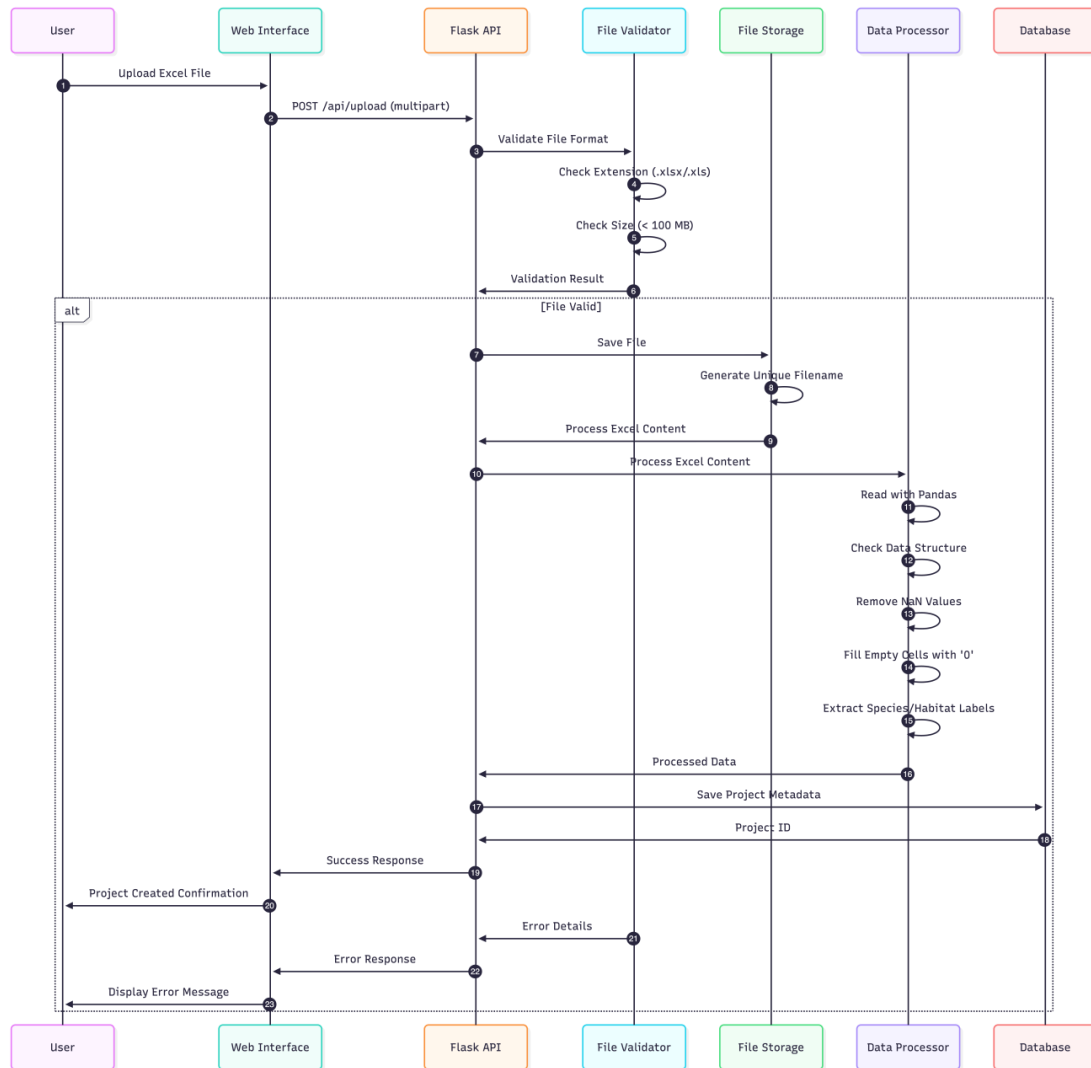


Fig. 13. Sequence diagram illustrating the data upload and validation process architecture.

B. Ablation: Vectorization of the Preference-Index Core

The preference-index computation (`calculate_ipd`) was originally a per-species loop containing a nested per-habitat loop. It has been replaced by a single vectorized NumPy computation over all species at once, verified to reproduce the output of the original implementation to floating-point precision (`np.allclose`) on both real datasets and against the existing regression test suite. Table IV reports the cost of the two implementations over the synthetic size sweep.

The two implementations scale differently rather than merely differing by a constant factor. The cost of the legacy implementation grows approximately in proportion to the species count, a hundredfold increase in I raising it 66-fold, whereas the vectorized implementation rises only 2.8-fold over the same range and is effectively flat at the matrix sizes encountered in practice. The speedup is consequently not a fixed figure but an increasing function of dataset size, which is the property that matters for the largest matrices the method is intended to serve.

TABLE IV. ABLATION OF THE VECTORIZED MATHEMATICAL CORE: COST OF THE PREFERENCE-INDEX COMPUTATION WITH AND WITHOUT VECTORIZATION (MEAN OF FIVE TIMED EXECUTIONS)

Species count (I)	Legacy scalar loop (ms)	Vectorized (ms)	Speedup
10	2.59	0.73	3.5×
50	9.54	0.85	11.3×
100	17.94	0.82	21.8×
500	86.05	1.43	60.1×
1000	171.62	2.01	85.6×

C. Ablation: Sensitivity of the Classification Thresholds

The four threshold targets are the only quantities in the pipeline that are set by convention rather than derived from the data, and the preference class assigned to a species depends on them. Each was therefore perturbed independently by ± 0.05 and ± 0.10 on the SM1 dataset, and the resulting change in each species' assigned class was measured against the default-threshold baseline (Table V). The reported case is habitat AA, in which seven species have non-zero abundance, and the baseline classification is three Preferential, three Transgressive, and one foreign species.

TABLE V. SENSITIVITY OF THE PREFERENCE CLASSIFICATION TO PERTURBATION OF THE FOUR THRESHOLD TARGETS (SM1, HABITAT AA, SEVEN SPECIES WITH NON-ZERO ABUNDANCE)

Threshold target	Perturbation	Species reclassified
0.1	+0.05, +0.10, -0.10	1 / 7 (14.3%)
0.1	-0.05	0 / 7
0.3	± 0.05 , ± 0.10	0 / 7
0.5	-0.05, -0.10	1 / 7 (14.3%)
0.5	+0.05, +0.10	0 / 7
0.75	± 0.05 , ± 0.10	0 / 7
All four (tolerance widened)	± 0.05 , ± 0.10	0 / 7

The 0.1 and 0.5 thresholds each independently drive a reclassification for this dataset, whereas 0.3 and 0.75 are stable across the tested range. The direction of the effect differs between them: at 0.1, three of the four tested perturbations reclassify a species, all except the smallest downward shift; at 0.5, only lowering the threshold reclassifies a species; raising it does not. That widening the search tolerance around all four targets simultaneously reclassifies no species — in contrast to the individual target-value shifts, which do — indicates the classification is sensitive to where each threshold's target value is set, but robust to the width of the tolerance band used to locate it on the exclusive species curve; these are two distinct senses of “sensitivity” worth keeping separate. The thresholds are, on this evidence, a defensible convention rather than a tuned parameter, though the analysis is reported for one habitat of one dataset and should not be read as a general stability result.

D. Scalability of the Full Pipeline

Table VI reports end-to-end pipeline latency on the two real datasets.

TABLE VI. FULL-PIPELINE LATENCY ON THE TWO REAL DATASETS (MEAN OF FIVE TIMED EXECUTIONS)

Dataset	Species (I)	Habitats (J)	ESC points [J/0.0015]	Full-pipeline latency
SM1 (benthic invertebrates, Sebou)	93	37	24,667	511 ms
SM2 (waterbird winter counts)	114	171	114,000	2.32 s

Once the preference-index core is vectorized, latency is dominated by the habitat count rather than by the species count. Between SM1 and SM2, the species count rises by a factor of only 1.2, while the habitat count rises by a factor of 4.6 and latency rises by a factor of 4.5 — closely tracking the habitat count rather than the species count. The explanation is that the remaining non-vectorized step, the construction of the abundance scale, carries a cost that scales with the $[J / 0.0015]$ evaluation points of the ESC, and therefore with J. This step was outside the scope of the present optimization and is reported here as a known limitation rather than left implicit: it is the natural target for further work, and the point at which additional effort would yield the largest return.

VII. DISCUSSION

Even though the APD method is forty years old, its use remained difficult, mainly because of the complex calculations that it involves. The algorithm published in [6] was a significant step towards implementation; however, given that very few ecologists are familiar with algorithm implementation, the use of this method still seemed difficult. The EcoAPD application was therefore a necessary step for ensuring the use of the APD method. To further facilitate this use in community ecology, the application has been made accessible to a wide audience, from any web browser, without installing any software.

From a scientific perspective, this application represents a significant advance in ecoinformatics [22], primarily as one of the rare tools designed to handle datasets in the form of contingency tables, given that such tables are widely used in ecology. Compared to traditional statistical approaches to analyzing these kinds of tables, EcoAPD offers another significant technical advancement, namely its ability to convert such tables into diagrams that provide a pedagogic visual representation of all species in a community, classifying them according to their degrees of preference. In addition, EcoAPD is available free of charge, giving a large number of scientists and biodiversity managers the opportunity to access a key method in ecology.

In summary, the EcoAPD application can address several key challenges in ecological informatics, mainly through:

- clear interface, which makes it easy to use with no need for specific informatics knowledge;
- accessibility, as it provides a web-based interface, eliminating barriers related to software installation and programming expertise;
- reproducibility, via its standardized implementation, ensuring the needs of transparency of the system across different studies [34], promoting a certain scientific rigor;

- data validation, as the system ensures data control, preventing errors that could compromise analytical results;
- flexibility, mainly due to the simultaneous treatment and display of normalized and non-normalized APDs, allowing users to select rapidly the APD version most suited to their data;
- scalability: vectorization of the preference-index core reduces its cost by a factor that rises with dataset size, reaching $85.6\times$ at 1000 species (Table IV), and the full pipeline processes the waterbird winter monitoring matrix in 2.32 s (Table VI), the residual cost being dominated by habitat count rather than species count (Section VI-D);
- data persistence, ensured by the use of a database-based mechanism that enables rapid output generation and visualization, which is especially advantageous for large datasets and iterative analysis workflows

VIII. CONCLUSION

The biggest gain from this research output lies in the facility and time saving offered by the EcoAPD application in implementing the APD method, through complex calculations performed behind the interface. Moreover, the output requests benefit from graphical tools that make it user-friendly, practically with no need for additional help. These tools could bring great help in developing similar applications.

Processing of the 114×171 waterbird matrix completes in 2.32 s (Table VI). Once the preference-index core is vectorized, full-pipeline latency is dominated by the habitat count rather than the species count; the remaining non-vectorized step — construction of the abundance scale, whose cost scales with $[J / 0.0015]$ — accounts for the greater part of the residual time. Further efficiency in calculation, storage and plotting therefore remains attainable, and the application can be improved in several ways:

- interface improvement: primarily by making better basic information on DP concept and the APD method, making the tool more educational and accessible;
- performance optimization: by vectorizing the construction of the abundance scale, which Section VI-D identifies as the dominant remaining cost, and by reducing the duration of the plotting process, particularly for datasets with many habitats;
- enhanced features: additional visualization options, statistical summaries, and export formats could further enhance the application's utility;
- integration capabilities: future versions could integrate other ecological analysis tools and databases, offering more comprehensive ecosystem analyses.

The immediate perspectives consist of improving the interface, in response to the ecologists' requirements. The main output expected from the use of the APD, however, is a reinterpretation of former assemblage studies that were based on abundances alone. In the expectation of a wide use of the APD

method in conservation and management decisions, the intention is to make it accessible to broader ecology and nature management communities.

SUPPLEMENTARY MATERIALS

The following supporting information can be downloaded at: 10.5281/zenodo.21504481. Table SM1: Matrix of data used to provide experimental examples for implementing the APD method; results of water insect surveys in the Upper Sebou River (Middle Atlas, Morocco). Table SM2: Matrix of data used to provide experimental examples for implementing the APD method; results of waterbird winter monitoring in Morocco in 2018-2022. Script SM3: Benchmark and ablation scripts, with their raw CSV outputs, underlying the evaluation reported in Section VI (Tables IV-VI).

REFERENCES

- [1] Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- [2] Hirzel, A. H., & Le Lay, G. (2008). Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, 45, 1372–1381. <https://doi.org/10.1111/j.1365-2664.2008.01524.x>
- [3] MacArthur, R. H. (1972). *Geographical ecology: Patterns in the distribution of species*. Harper & Row Publishers.
- [4] Gilbert, N. A., Amaral, B. R., Smith, O. M., & others. (2024). A century of statistical ecology. *Ecology*, 105, e4283. <https://doi.org/10.1002/ecy.4283>
- [5] Dakki, M. (1987). *Running Water Ecosystem of the Upper Sebou (Middle Atlas, Morocco) (in French) (Vol. 42)*.
- [6] Dakki, M., Ennakri, M., El Farouki, M., Ziti, S., & Zoglat, A. (2025). A Graphical Method for Analysing Species Assemblages Combining Habitat Preference and Abundance. *African Journal of Ecology*, 63, 1–13. <https://doi.org/10.1111/aje.70013>
- [7] Ennakri, M., Dakki, M., El Farouki, M., Ziti, S., & Zoglat, A. (2025). The Abundance-Preference Method for Assemblage Analysis: A Normalized Diagram with Algorithm Implementing the Method. *Diversity*, 17(9), 598. <https://doi.org/10.3390/d17090598>
- [8] Barraquand, F., Ezard, T. H., Jørgensen, P. S., & others. (2014). Lack of quantitative training among early-career ecologists: A survey of the problem and potential solutions. *PeerJ*, 2, e285. <https://doi.org/10.7717/peerj.285>
- [9] Vitolo, C., Elkhatib, Y., Reusser, D., Macleod, C. J. A., & Buytaert, W. (2015). Web technologies for environmental big data. *Environmental Modelling & Software*, 63, 185–198. <https://doi.org/10.1016/j.envsoft.2014.10.007>
- [10] Hirzel, A. H., Hausser, J., Chessel, D., & Perrin, N. (2007). *Biomapper 4.0 [Computer software]*. Laboratory for Conservation Biology, University of Lausanne.
- [11] Hirzel, A. H., Hausser, J., Chessel, D., & Perrin, N. (2002). Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*, 83(7), 2027–2036. [https://doi.org/10.1890/0012-9658\(2002\)083\[2027:ENFAHT\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2027:ENFAHT]2.0.CO;2)
- [12] Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3), 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- [13] Oksanen, J., Simpson, G. L., Blanchet, F. G., & others. (2022). *vegan: Community Ecology Package, Version 2.6-4*. R Foundation for Statistical Computing.
- [14] Mengin, N., Thorel, A., Liebig, H., & Segura, G. (2002). *ProCURVE Technical Note: Software for Calculating Habitat Preferences for Aquatic Organisms* | *International Journal of Limnology – Annales de Limnologie (in French) - International Journal of Limnology*, 38(4), 329–338. <https://doi.org/10.1051/limn/2002027>

- [15] Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, 22(4). <https://doi.org/10.18637/jss.v022.i04>
- [16] McCune, B., M. J., & Mefford. (1998). PC-ORD: Multivariate Analysis of Ecological Data. *The Bulletin of the Ecological Society of America*, 79(2), 144–145. <https://doi.org/10.2307/20168234>
- [17] Ter Braak, C. J. F., & Šmilauer, P. (2012). *Canoco Reference Manual and User's Guide: Software for Ordination, Version 5.0*. Microcomputer Power.
- [18] Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, 7, 1451–1456. <https://doi.org/10.1111/2041-210X.12613>
- [19] Edwards, J. L., Lane, M. A., & Nielsen, E. S. (2000). Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop. *Science*, 289(5488), 2312–2314. <https://doi.org/10.1126/science.289.5488.2312>
- [20] Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: Supporting ecology as a data-intensive science. *Trends in Ecology & Evolution*, 27(2), 85–93. <https://doi.org/10.1016/j.tree.2011.11.016>
- [21] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- [22] Hill, M. O. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, 54(2), 427–432. <https://doi.org/10.2307/1934352>
- [23] Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.
- [24] World Wide Web Consortium. (2014). *HTML5: A Vocabulary and Associated APIs for HTML and XHTML*.
- [25] World Wide Web Consortium. (2011). *Cascading Style Sheets (CSS)*
- [26] Ecma International. (2023). *ECMAScript® 2023 Language Specification*.
- [27] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- [28] McKinney, W. (2010). Data Structures for Statistical Computing in Python—SciPy Proceedings. <https://proceedings.scipy.org/articles/Majora-92bf1922-00a>
- [29] The pandas development team. (2024). *pandas-dev/pandas: Pandas (Version v2.2.3) [Computer software]*. Zenodo. <https://doi.org/10.5281/zenodo.3509134>
- [30] Oracle Corporation. (2023). *MySQL 8.0 Reference Manual*.
- [31] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9, 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- [32] Li, D., Mei, H., Shen, Y., Su, S., Zhang, W., Wang, J., Zu, M., & Chen, W. (2018). ECharts: A declarative framework for rapid construction of web-based visualization. *Visual Informatics*, 2(2), 136–146. <https://doi.org/10.1016/j.visinf.2018.04.011>
- [33] von Ahn, L., Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science*, 321(5895), 1465–1468. <https://doi.org/10.1126/science.1160379>
- [34] Powers, S. M., & Hampton, S. E. (2019). Open science, reproducibility, and transparency in ecology. *Ecological Applications*, 29(1), e01822. <https://doi.org/10.1002/eap.1822>