

Методика создания набора данных точечных карт находок в монографии «Сосудистые растения советского Дальнего Востока»

Methodology for creating a dataset of point maps of finds in the monograph “Vascular Plants of the Soviet Far East”

Поспелов И. Н.¹, Серегин А. П.²

Pospelov I. N.¹, Seregin A. P.²

¹ Институт проблем экологии и эволюции им. А. Н. Северцова РАН, г. Москва, Россия. E-mail: pleuropogon@gmail.com
¹ A. N. Severtsov Institute of Ecology and Evolution of the Russian Academy of Sciences, Moscow, Russia

² Московский государственный университет им. М. В. Ломоносова, биологический ф-т., г. Москва, Россия
E-mail: botanik.seregin@gmail.com
² Lomonosov Moscow State University. Faculty of Biology, Moscow, Russia

Реферат. Переведены в цифровой вид карты распространения высших сосудистых растений из девятитомной монографии «Сосудистые растения советского Дальнего Востока» (1985–1996, 2006). Исходные растровые карты привязывались вручную, после чего в ГИС ArcGIS создавался точечный слой находок растений. Всего было оцифровано 4387 карт, на которых было отмечено 156767 находок 4140 видов растений. При привязке точек находок в ряде случаев проводилась сверка с текстом аннотации распространения вида, могли добавляться не указанные на картах точки, или, напротив, не указываться ошибочные. Подробно излагается техническая методика привязки и оцифровки карт и основные возникавшие при этом проблемы. Проведена приблизительная оценка географической точности данных, составляющая в зависимости от масштаба и качества исходной карты от 1 до 25 км. Полученный набор данных был приведен к стандартам структуры и таксономической номенклатуры GBIF (Darvin Core) и опубликован там для общего доступа. Результаты существенно дополняют имеющиеся цифровые географические данные по флористическому разнообразию на территорию России, наряду с ранее опубликованными в GBIF данными аналогичного издания «Флора Сибири».

Ключевые слова. База данных, география растений, ГИС, карты распространения растений, Российский Дальний Восток, флора сосудистых растений, GBIF.

Summary. The maps of vascular plants distribution from the 9-volume monograph “Vascular Plants of the Soviet Far East” (1985–1996, 2006) are digitizing. The original raster maps were manually georeferenced and then a point layer of plant finds was created in ArcGIS. Totally, 4387 maps were digitized, recording 156767 finds of 4140 plant species. In some cases, a check was made against the text of the species distribution annotation when referencing the find points. In doing so, points not indicated on the maps may have been added, or, on the contrary, erroneous points may not have been indicated. The technical methodology of map georeferencing and digitization and the main problems encountered are described in detail. A rough estimate of the geographical accuracy of the data has been made, ranging from 1 to 25 km, depending on the scale and quality of the original map. The resulting dataset has been adapted to the GBIF structure and taxonomic nomenclature standards (Darvin Core) and published there for public access. The results add significantly to the available numerical geographical data on floral diversity within Russia, along with previously published GBIF data from a similar publication “Flora of Siberia”.

Key words. Database, GBIF, GIS, plant distribution maps, plant geography, Russian Far East, vascular flora.

Введение. Монография – определитель «Сосудистые растения советского Дальнего Востока» (далее СРДВ) является одним из крупнейших источников флористических данных для территории России. Она издавалась с 1985 по 1996 г., также в 2006 г. был издан том дополнений и изменений (далее для краткости – том 9). Всего в определителе приведена информация о распространении, морфологии и экологии 4387 видов высших сосудистых растений огромного региона, охватывающего природные зоны от арктических тундр и полярных пустынь до муссонных субтропиков. Это составляет около 1/3 от флоры России в целом. Для 4140 видов приводятся точечные карты распространения, перевод которых в цифровой набор данных и является предметом данной работы.

В территорию, охватываемую изданием, входят 8 современных регионов России – Приморский, Хабаровский и Камчатский края, Сахалинская, Амурская и Магаданская области, Чукотский автономный округ и Еврейская автономная область (3 112 700 км², 18,2 % от общей сухопутной площади РФ). На момент начала издания эта территория рассматривалась как Дальневосточный экономический район СССР.

Материалы и методы. Все карты распространения растений (1127 карт, табл. 1) из определителя были отсканированы. В аналогичном издании «Флора Сибири» соблюдался принцип «1 карта – 1 вид», а также число использованных «бланковок» различного охвата и масштаба было ограничено 10–15. Это позволило создать соответствующий набор данных, опубликованный в GBIF (Artemov, Egorova, 2021), в полуавтоматическом режиме. В СРДВ применялся другой подход. Во-первых, число видов на одной карте составляло от 1 до 8, показанных разными символами. Во-вторых, число использованных «бланковок» не поддается учету, фактически каждая карта имеет свои особенности охвата, например, с карт на всю территорию региона мог быть «срезан» крайний юг, север, запад и т.д. Проекция используемых карт прямо в издании не указывается, но, по всей видимости, это коническая проекция, широко используемая в контурных учебных картах на территорию СССР (за исключением наиболее «крупномасштабных» карт на юг Приморского края, где, вероятно, использованы рисованные вручную схемы). Географическая привязка осуществлялась вручную в программном пакете ArcGIS 10.8 интерактивно к векторной основе базового масштаба 1 : 6 000 000 и крупнее. Для привязки были использованы (в порядке значимости) береговая линия океана, линейные объекты гидросети (реки), полигональные объекты гидросети (озера), координатная сетка, границы регионов РФ; все использованные векторные слои взяты из открытых источников. Фрагмент привязанной карты приведен на рисунке. При привязке наибольшее внимание уделялось «значащей» части карты (например, на картах на весь регион до 1/3 приходится на территорию Якутии), естественно, что наибольшего соответствия оператор добивался на собственно регион, не обращая внимания на искажения, возникающие вне участка точек находок. Основным показателем качества привязки было максимальное совпадение карты с гидросетью и морским побережьем, т.к. точки на исходные карты скорее всего наносились именно с ориентировкой на гидросеть.

Таблица 1

Сводная таблица материалов монографии «Сосудистые растения советского Дальнего Востока»

Том	Видов в томе*	Карт	Видов на картах	Точек
1	488	162	468	16855
2	529	145	513	23437
3	637	121	622	21294
4	469	112	434	17026
5	477	133	475	21123
6	542	99	522	14540
7	488	105	488	22468
8	508	105	479	19461
доп. (9)	249**	145	145	563
Итого	4387	1127	4140***	156767

Примеч.: * Число видов в томе бралось из аннотации к соответствующему тому (не пересчитывалось вручную), оно может несколько отличаться, например, в 7 томе видов больше, чем 488, так как в тексте есть несколько видов, для которых карт в томе нет.

** В дополнительном томе 249 только новых для региона видов, всего же в нем представлена информация по 1352 видам (новые находки, хромосомные числа, изменения таксономии и др.).

*** 6 видов из более ранних томов дублированы на картах в 9 томе (*Tephrosia subdentata* (Bunge) Holub, *T. kawakamii* (Makino) Holub, *Epimedium koreanum* Nakai, *Hyperzia arctica* (Tolm.) Sipliv., *H. petrovii* Sipliv., *Orostachys aggregata* (Makino) H. Nara, поэтому число видов меньше арифметической суммы (4146).

Для привязки растров использовалась равновеликая коническая проекция Альберса со следующими параметрами: false_easting: 23500000,0; false_northing: 0,0; central_meridian: 135,0; standard_

parallel_1: 52,0; standard_parallel_2: 64,0; latitude_of_origin: 0,0; Linear Unit: Meter (1,0); Geographic Coordinate System: GCS_WGS_1984; Angular Unit: Degree (0,0174532925199433); Prime Meridian: Greenwich (0,0); Datum: D_WGS_1984.

В этой же проекции в формате .geotiff сохранялись привязанные карты, ссылки на которые доступны в наборе данных. При привязке в основном использованы алгоритмы трансформации посредством полинома 1 и 2 порядка, число точек совмещения составляло 10–30, общая средняя квадратическая ошибка по набору точек составляла 8–15. В процессе работы были выявлены наиболее характерные точки совмещения (привязки), и ее выполнение значительно ускорилось с повышением качества. Ряд карт имеет врезки, наиболее часто карта юга региона сопровождалась врезкой на Курильские острова; также отдельно иногда даются врезки на остров Врангеля и другие (даже на остров Монерон), но часто они, присутствуя на карте, не несли нагрузки. В ряде случаев врезка привязывалась отдельно (остров Врангеля), в этом случае отдельный растр врезки имеет индекс «а» (карта – 68, врезка – 68а). Врезка на Курильские острова отдельно не привязывалась, точки с нее переносились на векторную карту глазомерно. Несколько карт, состоящих из 2 равноценных частей (Сахалин и Курильские о-ва), привязывались отдельно, соответственно растры имеют индексы «а» и «б».

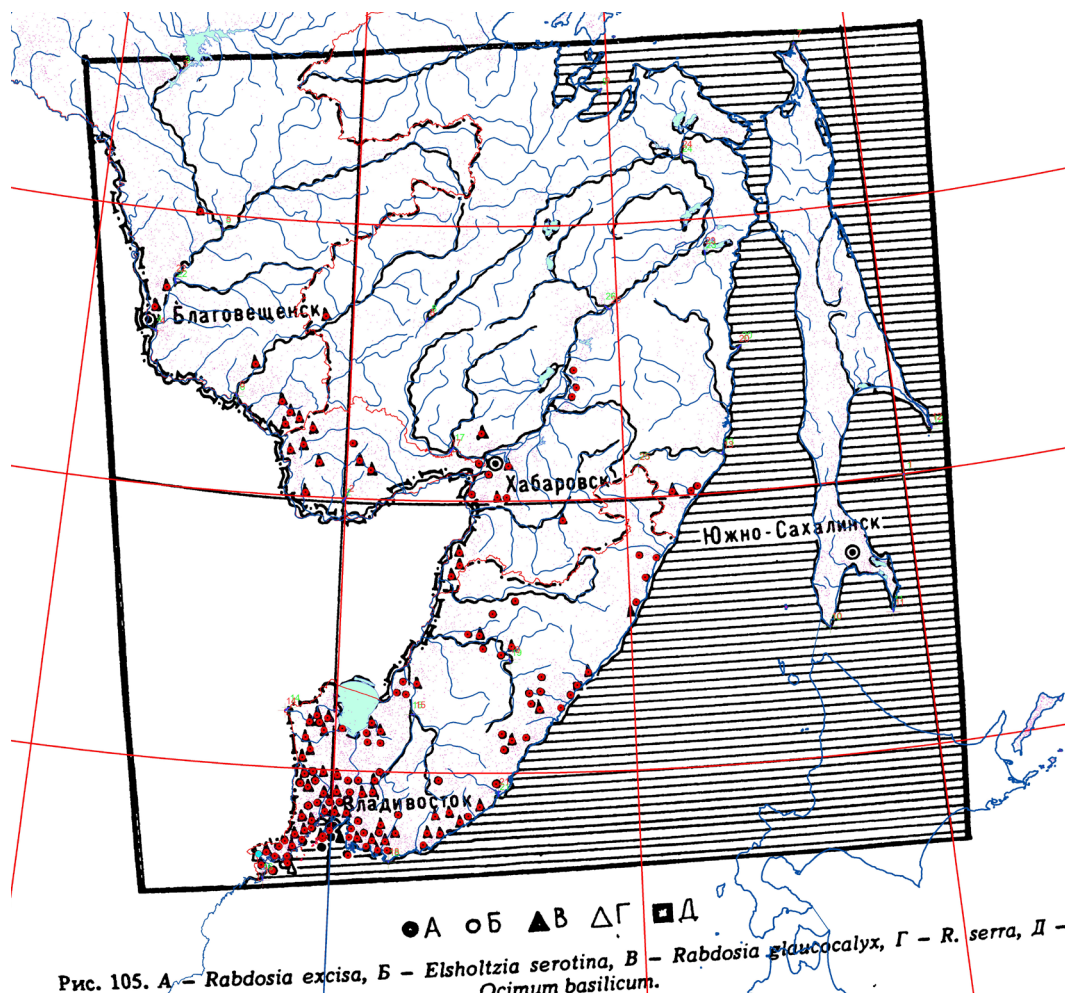


Рис. Фрагмент привязанной карты из СРДВ с нанесенными точками находок.

Результаты. По опыту оцифровки приведена примерная оценки точности данных. Размер одного символа на исходной карте составлял 1–1,2 мм. Для оценки точности привязок были выборочно измерены масштабы карт из «бумажного» издания. Выяснилось, что все карты можно условно поделить на 3 масштабные группы:

1. Карты на весь или почти весь регион. Масштаб: в 1 см = 220–250 км (1 : 22000000 – 1 : 25000000), размер символа на карте ок. 25 км, соответственно радиус 12,5 км, удвоен на компенсацию ошибок привязки растра и самого картографа – итого радиус точности устанавливался в 25 км.

2. Карты на $\frac{1}{4}$ – $\frac{2}{3}$ региона. Масштаб: в 1 см = 120–150 км (1:12 000 000 – 1:15 000 000), размер символа на карте ок. 12,5 км, – итого радиус точности устанавливался в 12,5 км.

3. Карты на небольшую часть региона (чаще всего юг Приморья) и врезки. Масштаб: в 1 см = 20–30 км (1:2 000 000 – 1:3 000 000), размер символа на карте ок. 5 км, – итого радиус точности устанавливался в 5 км. Вообще, на таких картах оператор опирался не столько на расположение точки на привязанном растре, сколько на пространственное отношение точки к указанным на карте географическим объектам. Для некоторых карт, вызвавших особенные сложности в привязке, точность выставлялась экспертно вне вышеназванных значений. Точки, попадающие на небольшие острова, имеют точность, соответствующую площади острова (особенно это заметно с о. Монерон, который на исходных картах просто отсутствовал, но нами эти точки устанавливались в контур острова).

После привязки каждой растровой карты точки находок вносились в точечный слой файловой базы геоданных ArGIS (отдельно для каждого тома). В первичную атрибутивную таблицу вносилось название вида, номер тома и карты, по окончании работы с каждым томом вычислялись координаты (lat/lon WGS 84) с помощью стандартного инструментария и добавлялось поле точности в соответствии с масштабом исходной карты (табл. 2). Точка, как правило, выставлялась в центроид символа на карте. Точки, расположенные в океане, перемещались на ближайшую сухопутную территорию (острова, побережья), также как и точки, выставленные за пределами границ России (как курьез нужно упомянуть, что на большинстве исходных карт к России «прирезали» заметный кусок Китая к северо-востоку от оз. Ханка). Напротив, несколько точек, поставленных на территорию республики Якутия, нами оставлены, т.к., по-видимому, они были там размещены авторами вполне сознательно.

Таблица 2

Фрагмент атрибутивной таблицы итогового набора данных

OBJECTID	Specie	volume	map	POINT_X	POINT_Y	acc_km
80	<i>Huperzia petrovii</i> Sipliv.	9	3	-178.90699	66.38182	25
81	<i>Huperzia petrovii</i> Sipliv.	9	3	-173.52551	64.42576	25
82	<i>Huperzia petrovii</i> Sipliv.	9	3	-172.50184	64.44476	25
83	<i>Huperzia petrovii</i> Sipliv.	9	3	-169.88874	66.08428	25
84	<i>Huperzia petrovii</i> Sipliv.	9	3	-178.19749	71.14055	25
85	<i>Diphasiastrum nikoense</i> (Franch. et Savat.) Holub	9	4	146.32077	44.36963	12.5
86	<i>Diphasiastrum nikoense</i> (Franch. et Savat.) Holub	9	4	148.38444	45.26704	12.5
87	<i>Diphasiastrum nikoense</i> (Franch. et Savat.) Holub	9	4	154.75946	49.32041	12.5
88	<i>Diphasiastrum nikoense</i> (Franch. et Savat.) Holub	9	4	154.85744	49.580621	12.5
89	<i>Diphasiastrum nikoense</i> (Franch. et Savat.) Holub	9	4	156.09095	50.71893	12.5
90	<i>Diphasiastrum nikoense</i> (Franch. et Savat.) Holub	9	4	156.79282	54.492974	12.5

Примеч.: данные в столбцах: OBJECTID – №№ записи в базе данных; Specie – название вида с патронимом; volume – том СРДВ; map – №№ рисунка карты в томе; POINT_X и POINT_Y – долгота и широта (WGS 1984, десятичные градусы); acc_km – оценочная точность находки (радиус в км).

Следует учитывать, что при составлении карт в СРДВ придерживались ряда общих правил «бумажной» картографии – а именно: недопущения перекрытия символов и установки символов на общегеографические объекты и их подписи (в данном случае города), этим объясняются «кольца» повышенной плотности точек вокруг Владивостока. Петропавловска-Камчатского и др. Особенную сложность в этом смысле представлял район залива Петра Великого, где в силу высокой плотности данных множество точек было выставлено в акваторию, причем часто в восточную часть залива, где островов практически нет; поэтому выбиралось просто линейно ближайшее побережье.

При оцифровке карт производилась сверка с текстом СРДВ. Часто возникали затруднения в правильной идентификации символов на «многосимвольных» картах (особенно трудно временами было различить символы «пустой круг» и «пустой квадрат»), а также трудностью нахождения отдельных одиночных символов на карте. При этом иногда обнаруживались несоответствия между текстом и картой, когда, например, символ есть на карте, но отсутствует в данном районе в тексте, и наоборот. В первом случае, символ оставался на карте; во втором – по возможности добавлялся на карту, если в тексте есть б.м. точная привязка (например – в т. 6 на карте 91 (*Taraxacum gorodkovii* Kharkev. et Tzvelev) – по тексту добавлена точка на Кроноцкий п-ов, отсутствующая на карте). Тем не менее ошибки, связанные с неправильной идентификацией символов в наборе данных, вполне вероятны, также за точки могли быть приняты типографские артефакты. Выявление и устранение подобных ошибок – задача последующей внимательной ревизии карт.

В томе 9, поскольку в нем в основном приводились новые для региона находки, и карты имеют в основном 1–5 точек, были привязаны только 19 карт из 145, точки с остальных карт наносились по аннотациям находок в тексте, координаты брались из картографических сервисов (Яндекс. Карты и др.), поэтому точность привязок там в основном выше.

Наименование таксона в атрибутивной таблице вводилось в строгом соответствии с таксономией, принятой в первоисточнике (даже в случае откровенно архаичных и невалидных названий), за исключением явных орфографических ошибок (*coraensis-coraensis*, *chancense-chancensis* и т.д.). Для этого все наименования таксонов сверялись (и копировались, во избежание опечаток) с базами данных флоры быв. СССР (по С. К. Черепанову (1995), а в ряде случаев – с сайтами Плантариум (<https://www.plantarium.ru>), The World Flora Online (<http://www.worldfloraonline.org>) и др. Это позволило минимизировать ошибки, хотя в патронимах они и весьма вероятны из-за неоднозначности их единого написания.

Полученный набор данных был приведен к стандартам структуры и таксономической номенклатуры GBIF (Darvin Core). Довольно немногочисленные таксономические и географические (попадание точек в океан) ошибки были сразу же исправлены, и массив данных на настоящий момент доступен для получения в GBIF (Pospelov, Seregin, 2023). Полученный в итоге набор данных содержит 156760 записей по 4140 видам растений (6 видов из более ранних томов дублированы на картах в 9 томе (*Tephrosia subdentata* (Bunge) Holub, *T. kawakamii* (Makino) Holub, *Epimedium koreanum* Nakai, *Huperzia arctica* (Tolm.) Sipliv., *H. petrovii* Sipliv., *Orostachys aggregata* (Makino) H. Hara, поэтому число видов меньше арифметической суммы видов по томам (4146). 427 видов представлены более чем 100 точками, из них 13 – более чем 300-ми (*Thalictrum minus* L. (491), *Galium boreale* L. (469), *Carex pallida* C. A. Mey. (401), *Parnassia palustris* L. (380), *Pedicularis resupinata* L. (363), *Festuca rubra* L. (338), *Salix schwerinii* E.L. Wolf (328), *Salix udensis* Trautv. et C. A. Mey. in Middend. (328), *Carex globularis* L. (327), *Lathyrus pilosus* Cham. (325), *Iris setosa* Pall. ex Link in Spreng., Schrad. et Lin (321), *Caltha palustris* L. (317), *Ribes triste* Pall. (310). Напротив, 871 вид имеют 1–3 точки находок (429 видов – 1 точка, 276 – 2 точки, 166 – 3 точки).

В дальнейшем планируется разместить для доступа на внешних серверах непосредственно привязанные растры карт находок с предоставлением в атрибутике набора данных ссылок на них, чтобы у каждого конкретного исследователя была возможность сверки оцифрованных данных с первоисточником. Также осуществляется географический анализ полученных данных, который послужит темой отдельных дальнейших публикаций.

ЛИТЕРАТУРА

- Плантариум. Растения и лишайники России и сопредельных стран: открытый онлайн атлас и определитель растений, 2007–2023. URL: <https://www.plantarium.ru/> (дата обращения: 08.05.2023).
- Сосудистые растения советского Дальнего Востока / Отв. ред. С. С. Харкевич. – Л.-СПб.: Наука, 1985–1996. – Т. 1–8.
- Флора российского Дальнего Востока: Дополнения и изменения к изданию «Сосудистые растения советского Дальнего Востока». Т. 9 / Отв. ред. А. Е. Кожевников и Н. С. Пробатова. – Владивосток: Наука, 2006. – 456 с.
- Черепанов С. К. Сосудистые растения России и сопредельных государств. – СПб.: Мир и Семья, 1995. – 990 с. Яндекс. Карты. URL: <https://yandex.ru/maps/> (дата обращения: 08.05.2023).
- Artemov I., Egorova A. Locations of plants on dot distribution maps in the Flora of Siberia (Flora Siberia, 1987–1997). Version 1.2. Central Siberian Botanical Garden SB RAS, 2021. Occurrence dataset. DOI: 10.15468/jb84wg (accessed via GBIF.org on 29 March 2023).

Pospelov I. N., Seregin A. P. (2023). Vascular plants of the Soviet Far East (1985–2006): georeferenced records from dot maps. Version 1.3. Lomonosov Moscow State University, 2023. Occurrence dataset. DOI: 10.15468/pa3dv8 (accessed via GBIF.org on 28 March 2023).

The World Flora Online. URL: <https://www.worldfloraonline.org> (Accessed 08 May 2023).