

Pelle László:

Az északi félgömb hőmérsékleti mezőjének
információtartalma

Doktori értekezés

Szeged
1981.

A meteorológia tudományának egyik fontos feladata az időjárás előrejelzése. Ez a nehéz és hálátlan feladat korántsem megoldott, bizonyos vizsgálatok azt mutatják, hogy tetszőleges időre és teljes pontossággal nem is valósítható meg. Becslések szerint a pontos numerikus előrejelzés kiterjesztésének elméleti határa kettő-három hét. Az ultrarövid, rövid és középtávú prognózisok matematikai, fizikai alapokon nyugvó módszere viszonylag fejlettek. Nem így a hosszútávúaké, ahol főként matematikai statisztikai eljárásokat használnak.

A következőkben bemutatandó módszer célja az, hogy segítséget nyújtson a hosszútávú előrejelzések elkészítéséhez. Mivel tisztán matematikai statisztikai jellégű, így korlátaira a kapott eredmények értékelésekor is tekintettel kell lennünk. Az elmúlt 100 - 150 év alatt végzett észlelések lehetőségét adnak egy konkrét hely éghajlatának jellemzéséhez az ún. klimatológiai normálértékek megadásával. Módszerünk lehetőséget nyújt a klimatológiai ismeretek alapján éghajlati prognózis javítására.

A klimatológiai prognózist a következő formában mondhatjuk ki: például Szeged januári középhő-

mérsékletének mediánja $\underline{m} = 1,0^{\circ}\text{C}$, tehát ha x -el jelöljük a következő január középhőmérsékletét akkor a

$$\begin{aligned} P / x < m / &= 0,5 & / 1 / \\ P / x \geq m / &= 0,5 \end{aligned}$$

legegyszerűbb előrejelzést adhatjuk.

Legyen ξ diszkrét valószínűségi változó, lehetséges értékei: $/ x_1, \dots, x_n /$ eloszlása: $/ p_1, \dots, p_n /$ ahol p_i az x_i esemény bekövetkezésének valószínűsége.

ξ , $H(\xi)$ -vel jelölt entrópiáját a következő formulával definiálhatjuk:

$$H(\xi) = - \sum_{i=1}^n p_i \log_2 p_i \quad / 2 /$$

Az entrópia mértékét bit-nek nevezzük, ha a logaritmus 2-es alapu. Egységnyi - 1 bit - az olyan diszkrét valószínűségi változó entrópiája, mely 2 lehetséges értékét egyenlő valószínűséggel veszi fel. Az így definiált entrópia az információelmélet egyik központi fogalma. E tudományág alapjait C. Shannon vetette meg a második világháború után, a hírközlés problémáinak megoldására, de mint minden jól megalapozott matematikai elméletnek, így ennek is alkalmazása egészen különböző területeken történik.

A / 2 /-ben adott definíció a diszkrét valószínűségi változó bizonytalanságának mértékéeként is felfogható, ezt támasztja alá az, hogy rögzített n esetén maximumát egyenletes eloszlásnál veszi fel.

Legyen ξ_1, ξ_2 két független diszkrét valószínűségi változó, ekkor

$$H(\xi_1, \xi_2) = H(\xi_1) + H(\xi_2) \quad / 3 /$$

ahol (ξ_1, ξ_2) a ξ_1 -ből és ξ_2 -ből alkotott két dimenziós diszkrét valószínűségi változó entrópiája. Tehát függetlenség esetén az entrópiák összeadódnak.

A következő bevezetésre kerülő fogalom - gyakorlati szempontból is rendkívül fontos - ugyanis két valószínűségi változó kapcsolatának szorosságát méri.

Legyen ξ, η két diszkrét valószínűségi változó $I(\xi, \eta)$ -val jelölve kölcsönös információtartalmukat:

$$I(\xi, \eta) = H(\xi) + H(\eta) - H(\xi, \eta) \quad / 4 /$$

Amint látható, függetlenség esetén $I(\xi, \eta) = 0$, a kölcsönös információtartalom szimmetrikus ξ -re η -ra, maximuma pedig a két entrópia közül a kisebbik: $\max(I(\xi, \eta)) = \min(H(\xi), H(\eta))$, ez az eset

akkor áll fenn, ha a két diszkrét valószínűségi változó függvénye egymásnak, tehát az egyik a másikkal kifejezhető. Természetesen bennünket ezen szélsőséges eseteknél sokkal inkább a valóságban gyakrabban előforduló sztohasztikus kapcsolatok érdekelnek, mivel a / 4 / mennyiséget azok erősségének mérésére fogjuk használni.

A / 2 / és / 4 / definíciók különösebb nehézség nélkül átvihetők folytonos valószínűségi változókra is.

Legyen ξ folytonos valószínűségi változó sűrűség függvénye $f(x)$ ekkor:

$$H(\xi) = - \int_{-\infty}^{+\infty} f(x) \log f(x) dx \quad / 5 /$$

A továbbiakban e alapu logaritmust használunk a számítások megkönnyítése érdekében, mivel ebből könnyen számíthatók az információmennyiségek bit-ben is.

Ha ξ és η folytonos valószínűségi változó, eloszlásfüggvényeik pedig $h(x)$, $g(y)$, együttes eloszlásfüggvényük $f(x,y)$ akkor:

$$I(\xi, \eta) = - \int h(x) \log h(x) dx - \int g(y) \log g(y) dy + \iint f(x,y) \log f(x,y) dx dy \quad / 6 /$$

A vizsgálatainkban előforduló havi átlag hőmérsékleti idősorokat normális eloszlású valószínűségi változóval írjuk le.

Ismert, hogy egy normális eloszlású valószínűségi változó sűrűségfüggvényét $\underline{m}$ várható értéke és szórása meghatározza:

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp \frac{(x-m)^2}{2 \sigma^2} \quad / 7 /$$

Ekkor / 5 / szerint ξ entrópiája

$$\begin{aligned} H(\xi) &= - \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi} \sigma} \exp - \frac{(x-m)^2}{2 \sigma^2} \log \frac{1}{\sqrt{2\pi} \sigma} \exp - \frac{(x-m)^2}{2 \sigma^2} dx = \\ &= - \log \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi} \sigma} \exp - \frac{(x-m)^2}{2 \sigma^2} dx + \frac{1}{2 \sigma^2} \int_{-\infty}^{+\infty} (x-m)^2 \cdot \frac{1}{\sqrt{2\pi} \sigma} \exp - \\ &\quad - \frac{(x-m)^2}{2 \sigma^2} dx = \log \sqrt{2\pi} \sigma + \log \sqrt{e} = \log \sqrt{2\pi e} \sigma \end{aligned}$$

Ez az eredmény azt mutatja, hogy az entrópia a jól ismert szórásfüggvénye, annak növekedésével ez is nő. Ha ξ , η normálisak, akkor ugyancsak / 5 / alapján: / együttes eloszlásfüggvényük $f(x,y)$ /

$$\begin{aligned} H(\xi, \eta) &= - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-r^2}} \exp \left[- \frac{1}{2(1-r^2)} \left(\frac{(x-m_x)^2}{\sigma_x^2} - 2r \frac{(x-m_x)(y-m_y)}{\sigma_x \sigma_y} + \right. \right. \\ &\quad \left. \left. + \frac{(y-m_y)^2}{\sigma_y^2} \right) \right] \cdot \log \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-r^2}} \exp \left[- \frac{1}{2(1-r^2)} \left(\frac{(x-m_x)^2}{\sigma_x^2} - 2r \frac{(x-m_x)(y-m_y)}{\sigma_x \sigma_y} + \right. \right. \\ &\quad \left. \left. + \frac{(y-m_y)^2}{\sigma_y^2} \right) \right] dx dy = - \log \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-r^2}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) dx dy + \frac{1}{2(1-r^2)} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \end{aligned}$$

$$\left(\frac{(x-m_x)^2}{\sigma_x^2} - 2r \frac{(x-m_x)(y-m_y)}{\sigma_x \sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2} \right) f(x, y) dx dy = \log 2\pi \sigma_x \sigma_y \sqrt{1-r^2} +$$

$$+ \frac{1}{2(1-r^2)} (2-2r) = \log 2\pi \sigma_x \sigma_y \sqrt{1-r^2} + 1 = \log 2\pi e \sigma_x \sigma_y \sqrt{1-r^2}$$

és így / 6 / szerint

$$I(\xi, \eta) = -\log \sqrt{2\pi} e \sigma_x - \log \sqrt{2\pi} e \sigma_y - \log 2\pi e \sigma_x \sigma_y \sqrt{1-r^2} = -\frac{1}{2} \log(1-r^2)$$

Tehát két normális valószínűségi változó kölcsönös információtartalma:

$$I(\xi, \eta) = -\frac{1}{2} \log(1-r^2)$$

ahol r a ξ és η közötti korrelációs együttható.
Ha ξ és η korrelálatlanok, akkor $I(\xi, \eta) = 0$,
ha $|r| \rightarrow 1$ akkor $I(\xi, \eta) \rightarrow \infty$ így elmondhatjuk,
hogy a $I(\xi, \eta)$ mennyiség a két valószínűségi változó kapcsolatának erősségét méri és szoros kapcsolatban áll a korrelációs együtthatóval.

A vizsgálat adatbázisa 83 az északi félgömbön elhelyezkedő állomás 1880-tól 1960-ig mért havi középhőmérsékletei. Sajnos az időszak elején még nem minden állomás működött és különböző események miatt később is keletkeztek hiányok.



1. Stykkisholm	N 64°30'	W 22°46'	14 m
2. Ponta Delgada	37 42	25 40	36
3. Funhal	32 37	16 54	56
4. Valencia	51 56	10 15	14
5. Aberdeen	57 10	2 6	58
6. Nantes	47 15	1 34	27
7. Madrid	40 24	3 41	660
8. Palma	39 33	E 2 42	43
9. Basel	47 33	7 35	317
10. De Bilt	52 6	5 11	2
11. Haparanda	65 50	24 9	7
12. Uppsala	59 51	17 38	15
13. Berlin	52 33	13 4	55
14. Róma	41 54	12 29	131
15. Lagos	6 24	3 24	38
16. Szeged	46 15	20 9	97
17. Arhangelszk	64 35	40 36	13
18. Moszkva	55 50	37 33	156
19. Kiev	50 27	30 30	179
20. Alexandria	31 12	29 53	7
21. Khartoum	15 37	32 33	389
22. Entebbe	0 5	32 29	1146
23. Orenburg	51 45	55 6	194
24. Tbiliszi	41 43	44 48	490
25. Aden	12 46	45 3	4

26. Bushire	N 29° 0'	E 59° 50'	14 m
27. Quetta	30 12	67 0	1320
28. Krasznovodsk	40 0	52 59	-20
29. Szurgut	61 15	73 30	43
30. Taskent	41 16	69 16	428
31. Barnaul	53 20	83 42	196
32. Almolinsk	51 12	71 23	347
33. Lahore	31 33	74 20	214
34. Bombay	18 54	72 49	11
35. Jaipur	26 55	75 50	436
36. Simla	31 6	77 10	2202
37. Allahabad	25 27	81 44	98
38. Nagpur	21 6	79 3	310
39. Bangalóre	12 58	77 35	921
40. Colombó	6 54	79 52	7
41. Madras	13 0	80 11	16
42. Calcutta	22 32	88 20	6
43. Akyab	20 8	92 53	5
44. Port Blair	11 41	92 45	20
45. Rangoon	16 46	96 10	23
46. Irkutszk	52 16	104 21	485
47. Neresinszk	51 19	119 37	621
48. Yakutszk	58 27	92 10	81
49. Nikolajerszk	53 9	140 42	47
50. Vladivosztok	43 7	131 54	138

51. Sapporó	N 43° 3'	E 141° 20'	18 m
52. Kyoto	35 1	135 44	43
53. Kagoshima	31 34	130 33	5
54. Mokpó	34 47	126 23	31
55. Honkong	22 18	114 10	33
56. Manila	14 35	120 59	16
57. Honolulu	21 18	W 158 6	3
58. Dawson	64 4	139 26	323
59. Dureka	40 48	124 10	24
60. Los Angeles	35 56	118 23	37
61. Barkerville	53 4	121 31	1274
62. Prince Albert	53 13	105 41	431
63. Havre	48 34	109 40	760
64. Mazatlan	20 58	89 38	22
65. Dodge City	37 46	99 58	796
66. Madison	43 8	89 20	265
67. Vicksburg	32 21	90 53	90
68. Merida	23 11	106 25	78
69. Torontó	43 40	79 24	116
70. Sandusky	41 27	82 43	192
71. Anticosti	49 24	63 33	7
72. Block Island	41 10	71 33	17
73. Lynchburg	37 20	79 12	291
74. Port au Prince	18 33	72 20	44
75. Bermuda	32 17	64 44	19

76. Trinidad	N 10°40'	W 61°31'	20 m
77. Georgetown	6 49	58 11	2
78. Upernavik	72 47	56 10	64
79. Jacobshavn	69 13	51 2	47
80. Godthaab	64 11	51 44	20
81. Ivigtut	61 12	48 10	30
82. Angmagalik	65 37	37 34	36
83. Petropovlovsk	52 59	E 158 39	24

A következőkben megvizsgáltuk, hogy az egyes havi hőmérsékleti mezők / északi félgömb / milyen információt szolgáltatnak egy kitüntetett állomás - Szeged - havi középhőmérsékleteire. Kiszámítottuk Szeged minden egyes hónapjának a teljes mező megelőző 12 hónapjával való kölcsönös információ-tartalmát. Pl. a szegedi október hőmérsékleti sorát leíró valószínűségi változónak és a teljes félgömb szeptemberi, augusztusi, júliusi, . . . , októberi mezőjét leírónak a kölcsönös információ-tartalmát. Így összesen $12 \times 12 = 144$ információ mennyiség mezőhöz jutottunk. Ezek mindegyike 83 db információ mennyiséget tartalmaz ez összesen 11952 számítást jelentett.

Ez igen munkaigényes feladat a kézi számolás gyakorlatilag kivihetetlen, a számításokat a JATE Kibernetikai Laboratóriumában végeztük el. Az 5 %-

os szignifikancia szinten elfogadható értékeket tekintettük elfogadhatónak / pl. $n = 80$ esetén

$I(\xi, \eta) = 0,0348$ bit-nek adódott /. Mint várható az információ mennyiségek döntő többsége ennél kisebb. De minden esetben több-kevesebb érték ennél nagyobb, a maximális kölcsönös információ mennyiségek $0,14 - 0,15$ bit körül vannak. A lehetséges 144 információ mennyiség térkép közül 14-et, az érdekesebbeket megrajzoltuk / 1. - 14. ábra /. Ezen a szignifikáns értékekkel rendelkező területek vastag vonallal határoltak. A " + " ill. " - " jelek a kapcsolatok irányára utalnak. A térképek alapján elmondható, hogy bizonyos területeknek kiemelt szerepe van különösen Észak-Amerika déli-keleti részének, Európa középső és nyugati térségeinek. Más területek Elő-India, Perzsa öböl, Mandzsúria - Kórea, Karib-szigetek csak egy-két esetben reprezentálnak jelentős kapcsolatokat.

Az információmennyiségeket két kategóriás valószínűségi előrejelzések készítésére használhatjuk, ugyanis a nyert információval arányosan az előrejelzés bizonytalansága csökken. Mivel a biztosan beváló kétkategóriás prognózishoz 1 bit információra van szükség, így a rendelkezésünkre álló $0,15$ bit maximális információmennyiség erre nem ad le-

hetőséget, viszont az / 1 / alatti klimatológiai prognózisnál jobbat készíthetünk. Ugyanis / 2 / szerint a kétkategóriás előrejelzés entrópiája

$$H = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$$

ennek maximuma $p = 0,5$, $1-p = 0,5$ esetén van $H_{\max} = 1$ bit, minden információ mennyiség nyereségével az entrópia csökken:

$$H = H_{\max} - I = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$$

így I ismeretében a / p , (1-p) / eloszlás meghatározható.

Pl.: Tekintsük konkrétan a februárt megelőző első hónapot itt a maximális információ mennyiség 0,148 bit, tehát

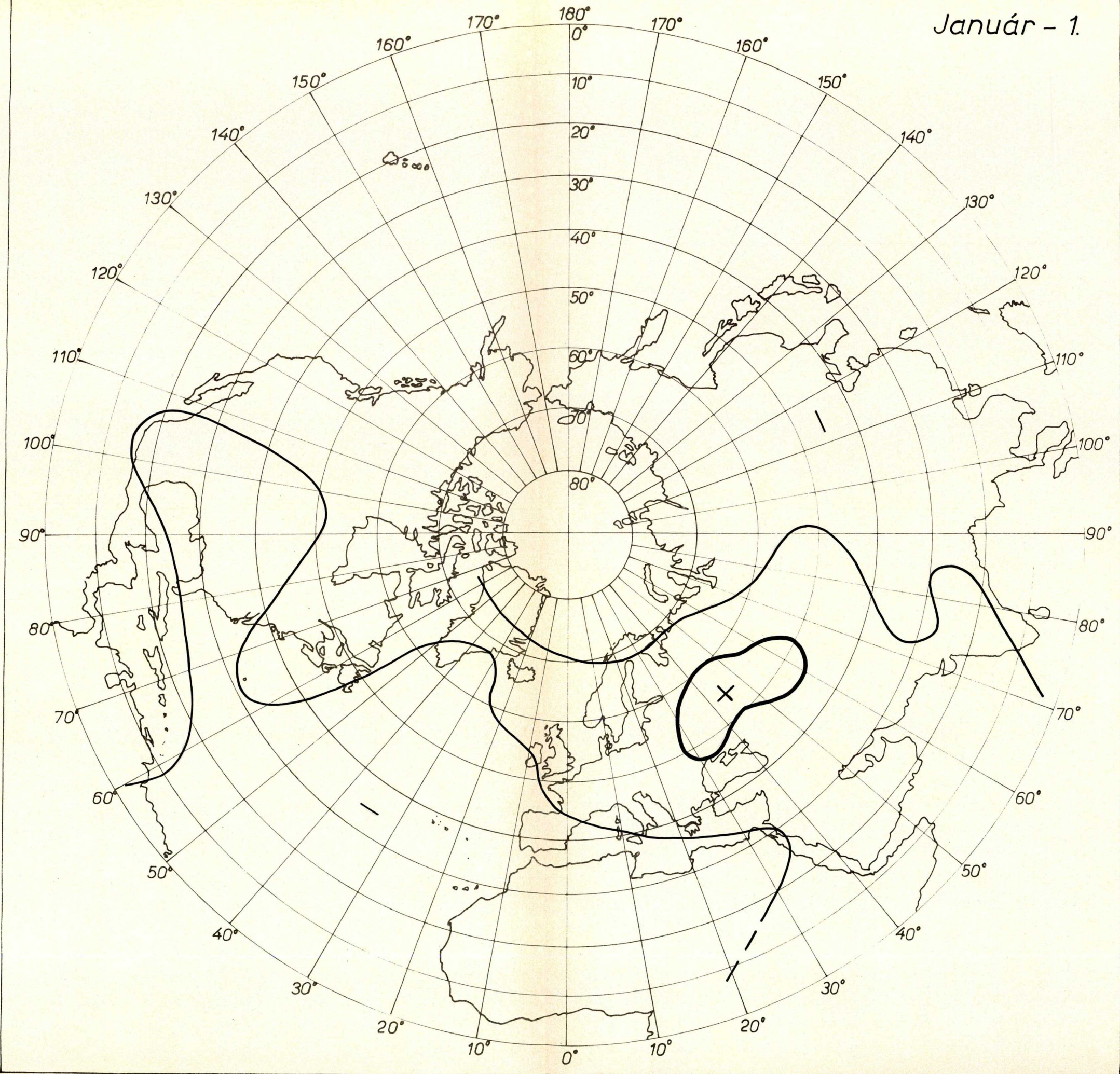
$$1 - 0,148 = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$$

ebből $p = 0,28$, $1-p = 0,72$ így a hőmérsékleti mező januári aktuális értékének ismeretében a szegedi februári középhőmérsékletre a / $p = 0,28$, $1-p = 0,72$ / előrejelzést adhatjuk és mivel a kapcsolat pozitív, így átlag alatti januári középhőmérséklet esetén az átlag alatti középhőmérsékletű február bekövetkezésének valószínűsége $p = 0,72$ és átlag feletti január esetén $p = 0,28$. Ezt a

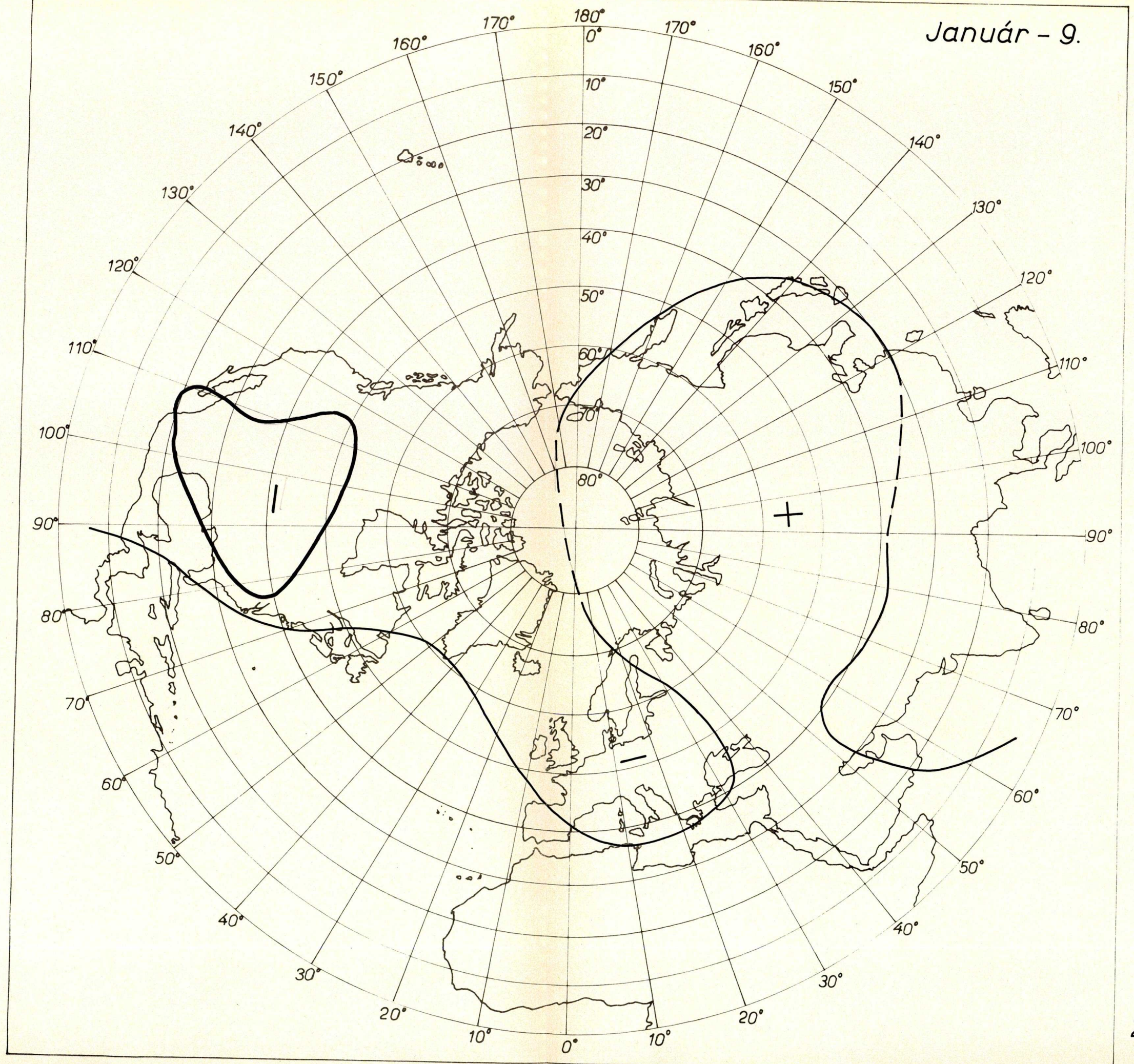
módszert alkalmaztuk az 1951 - 60-as évekre " archív " prognózisunk ekkor 65 %-an vált be, az 50 %-os éghajlati előrejelzéssel szemben. Előrejelzésünk alacsony hatékonysága érthető, mivel nagyon egyszerű prediktort használtunk fel, viszont jelenleg a jóval bonyolultabb módszerekkel készített, sokkal számítás igényesebb eljárások sem adnak lényegesen jobb eredményt.

I r o d a l o m

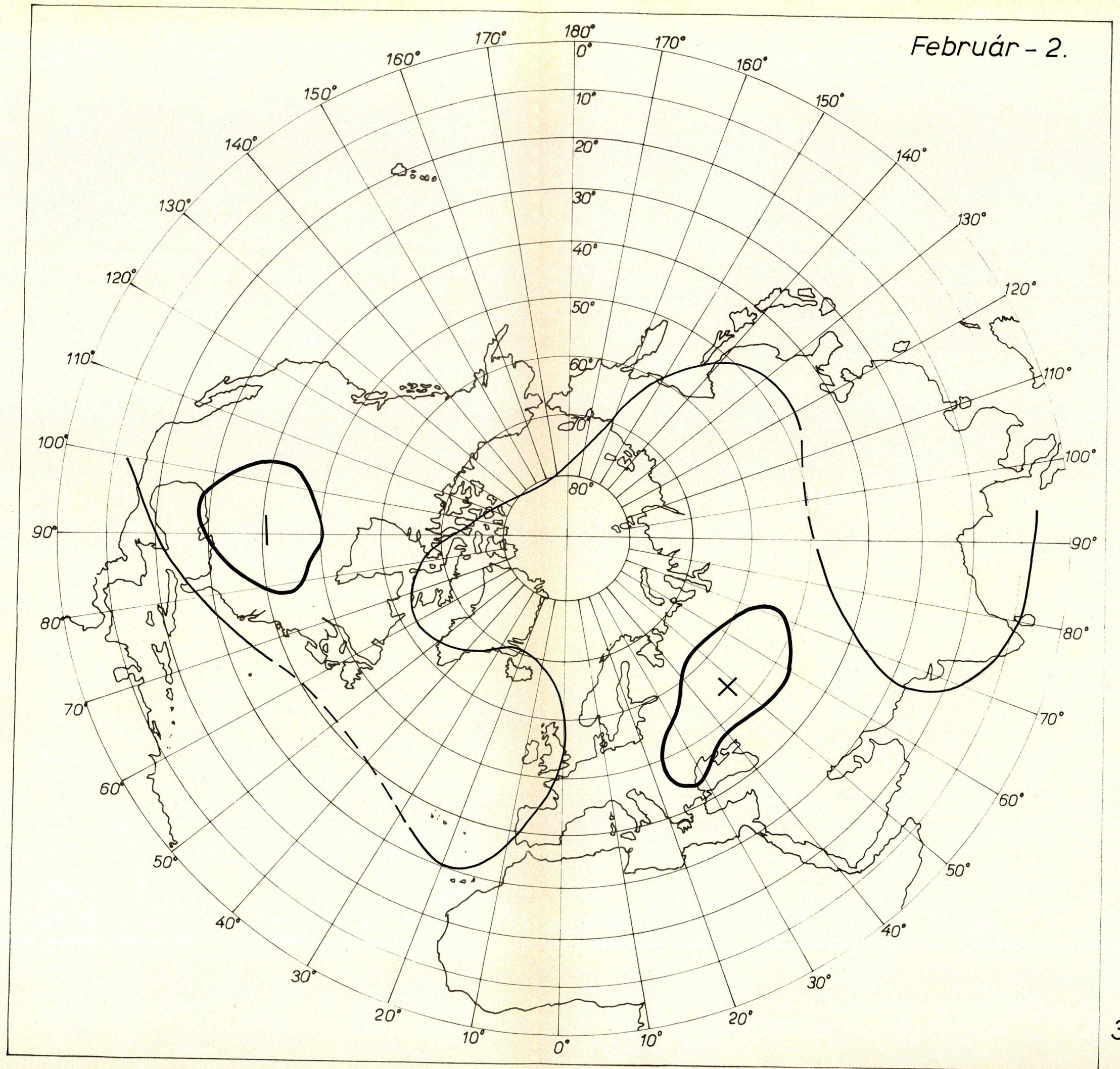
- 1 Péczely Gy.: Éghajlattan, jegyzet JATE TTK
Szeged 1977.
- 2 Rákóczi F. - Sziderovszky F.: Budapest hőmér-
sékleti sorának információ tar-
talma
Időjárás 80. évf./2.
- 3 Fazlollah M.Reza: Bevezetés az információ el-
méletébe
MK Budapest 1966.
- 4 World Weather Records, Wasinghton 1966.
- 5 Béll B.: Ciklusok és ritmusok a légköri moz-
gásrendszerekben
MTA X. Osztályának Közleményei
8/3-4. 1975.
- 6 Craddock J.M.: The Interannual Variability of
Monthly Mean Air Temperatures o-
ver the Northern Hemisphere
British Meteorological Office
Scientific Paper No. 20.



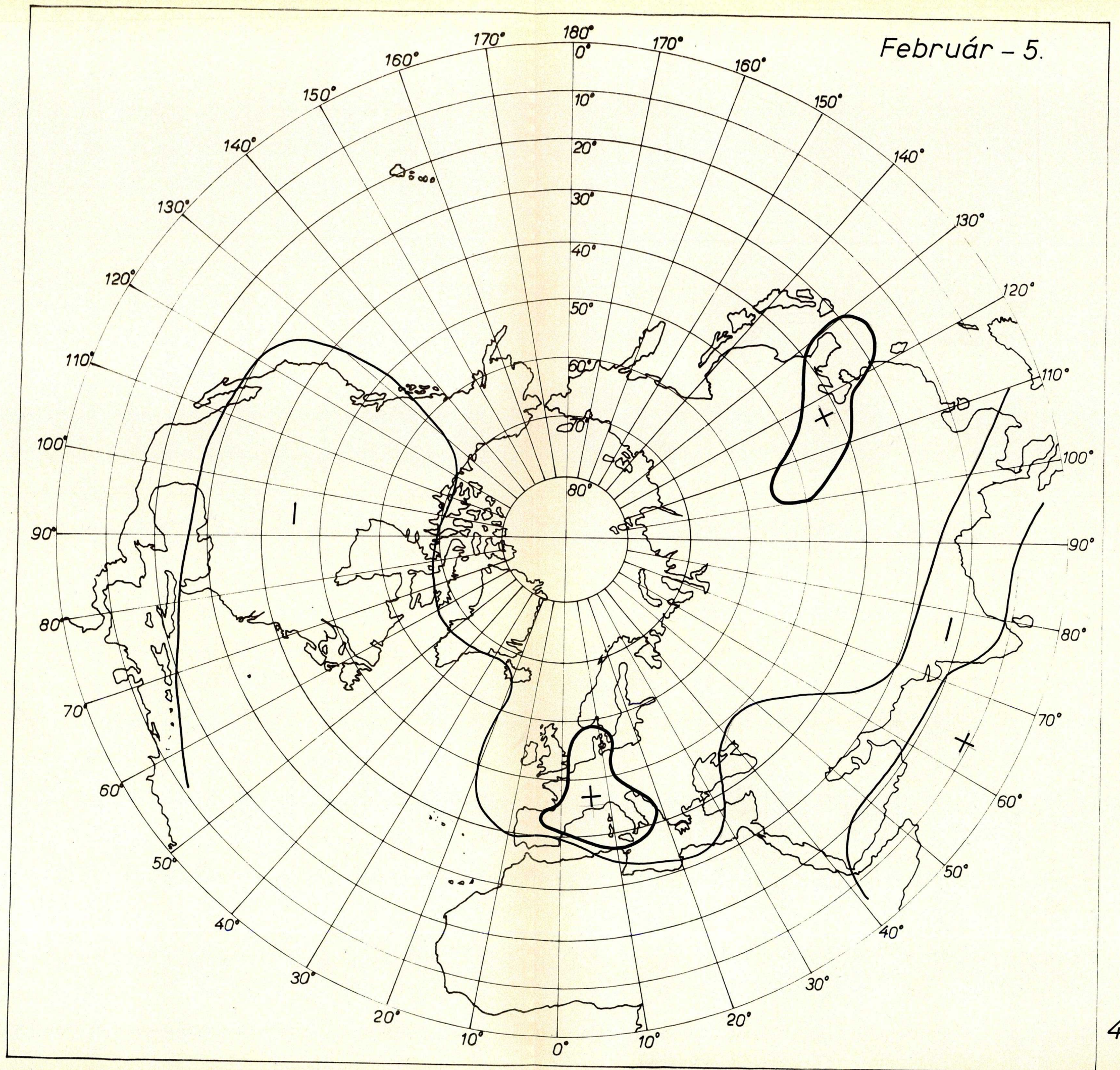
Január - 9.



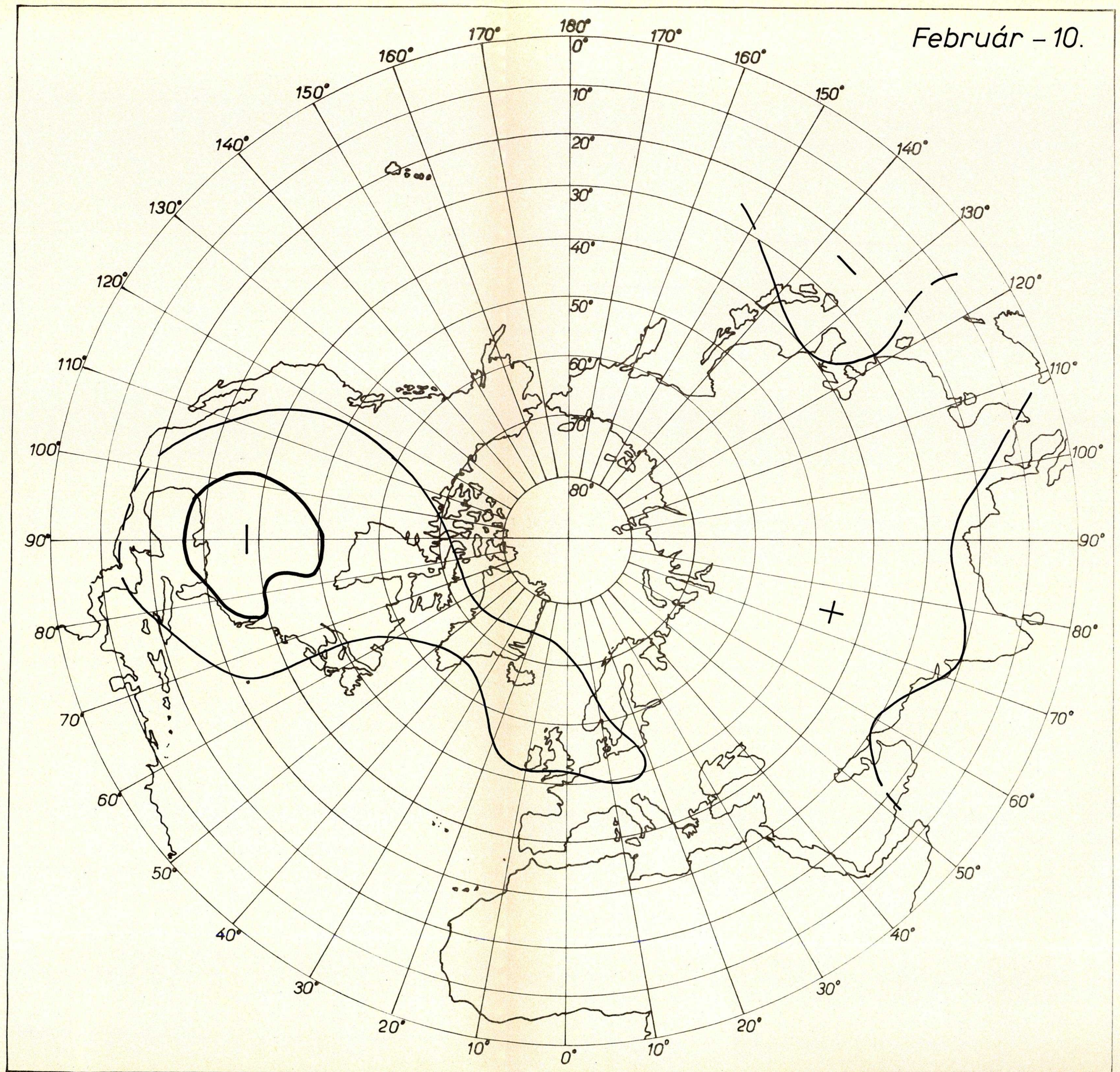
Február - 2.

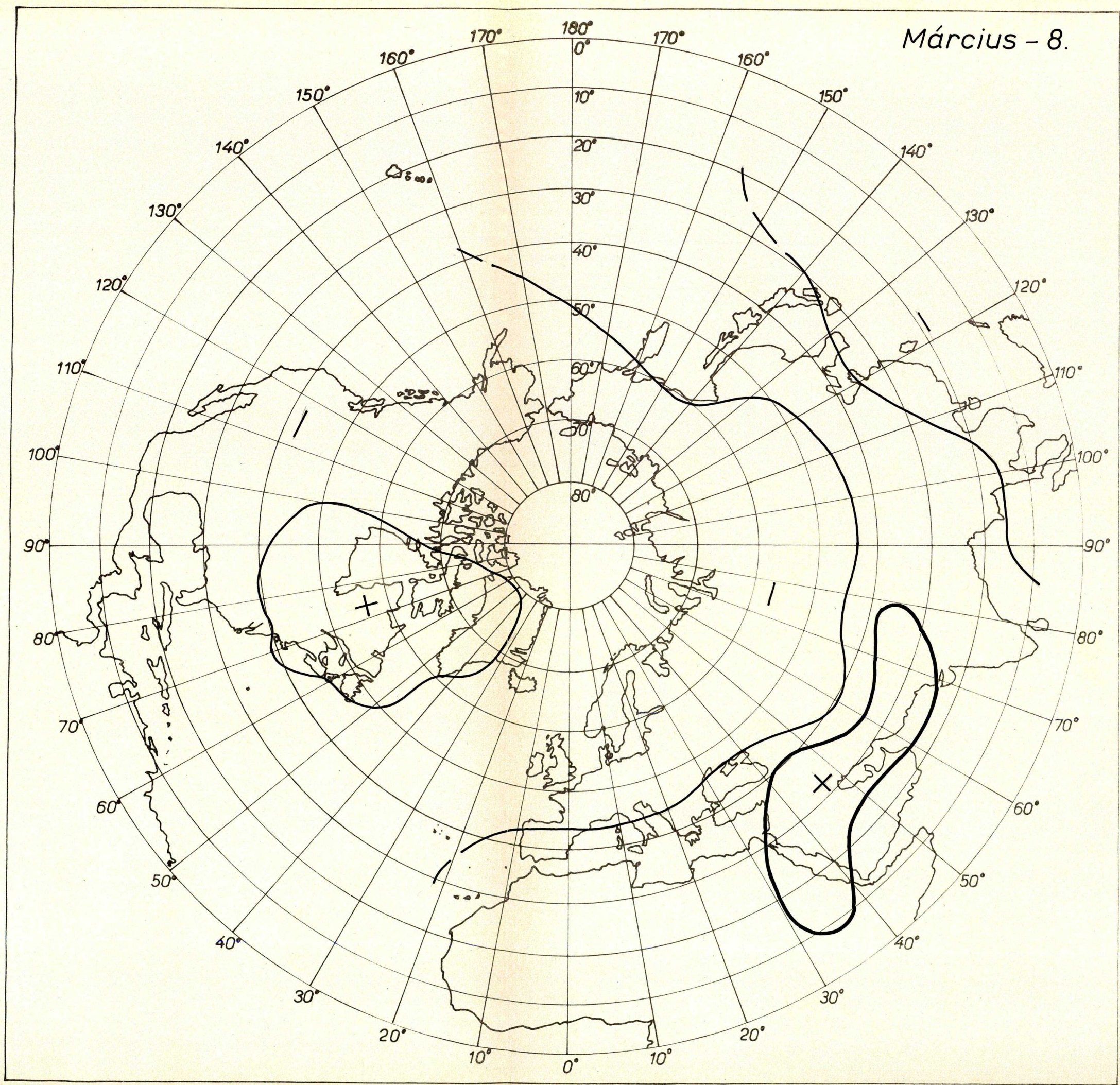


Február - 5.

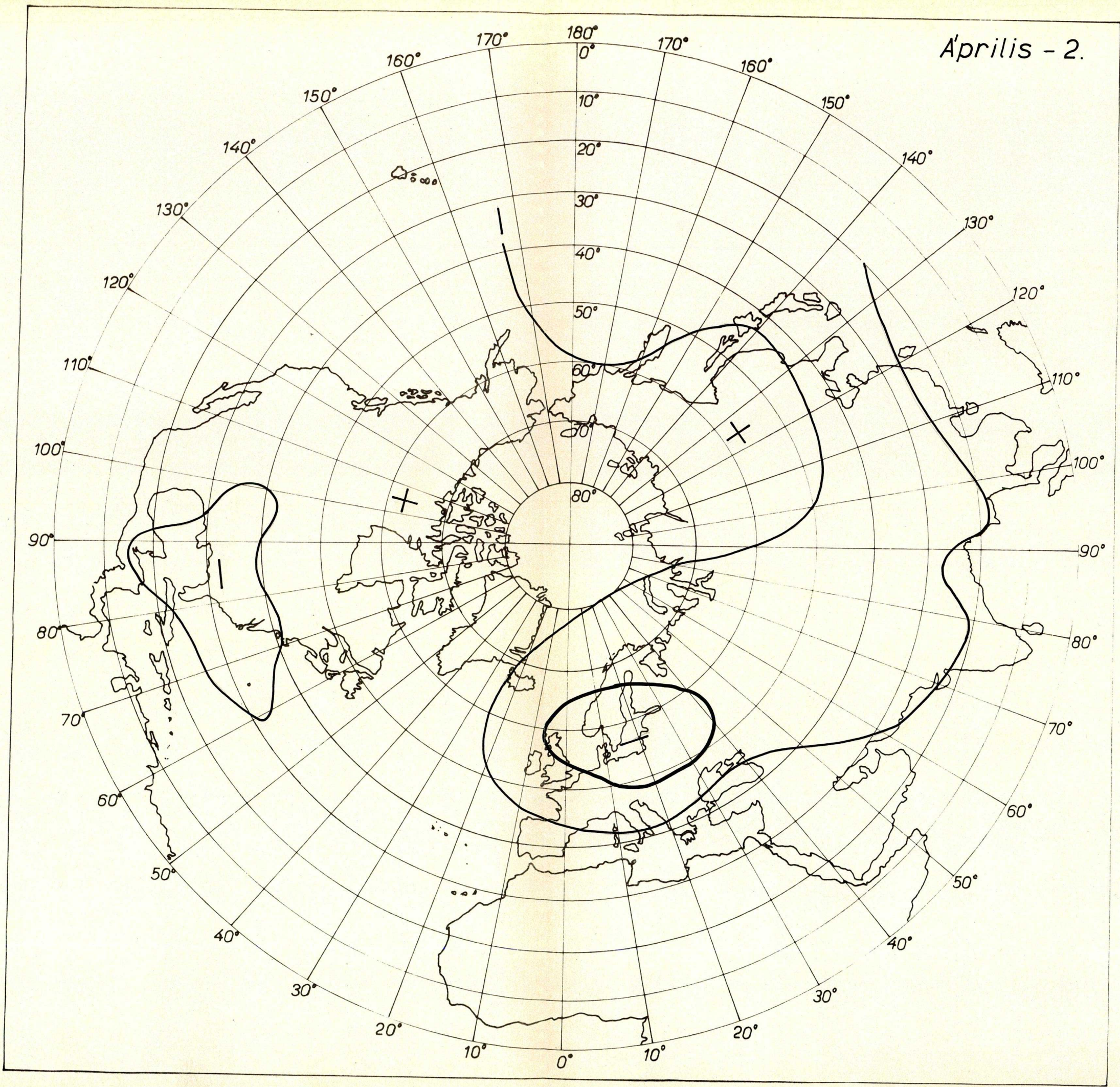


Február - 10.

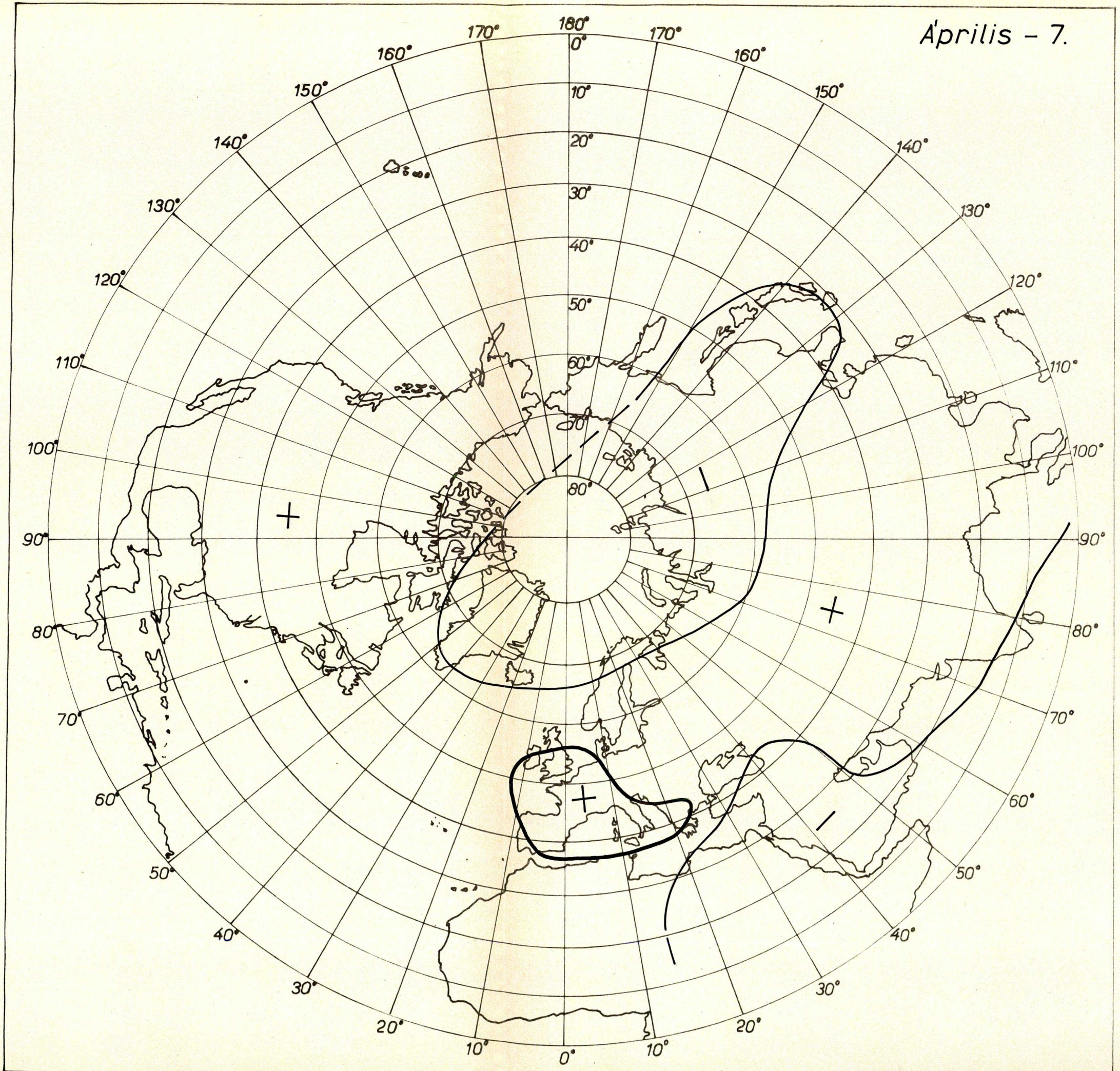




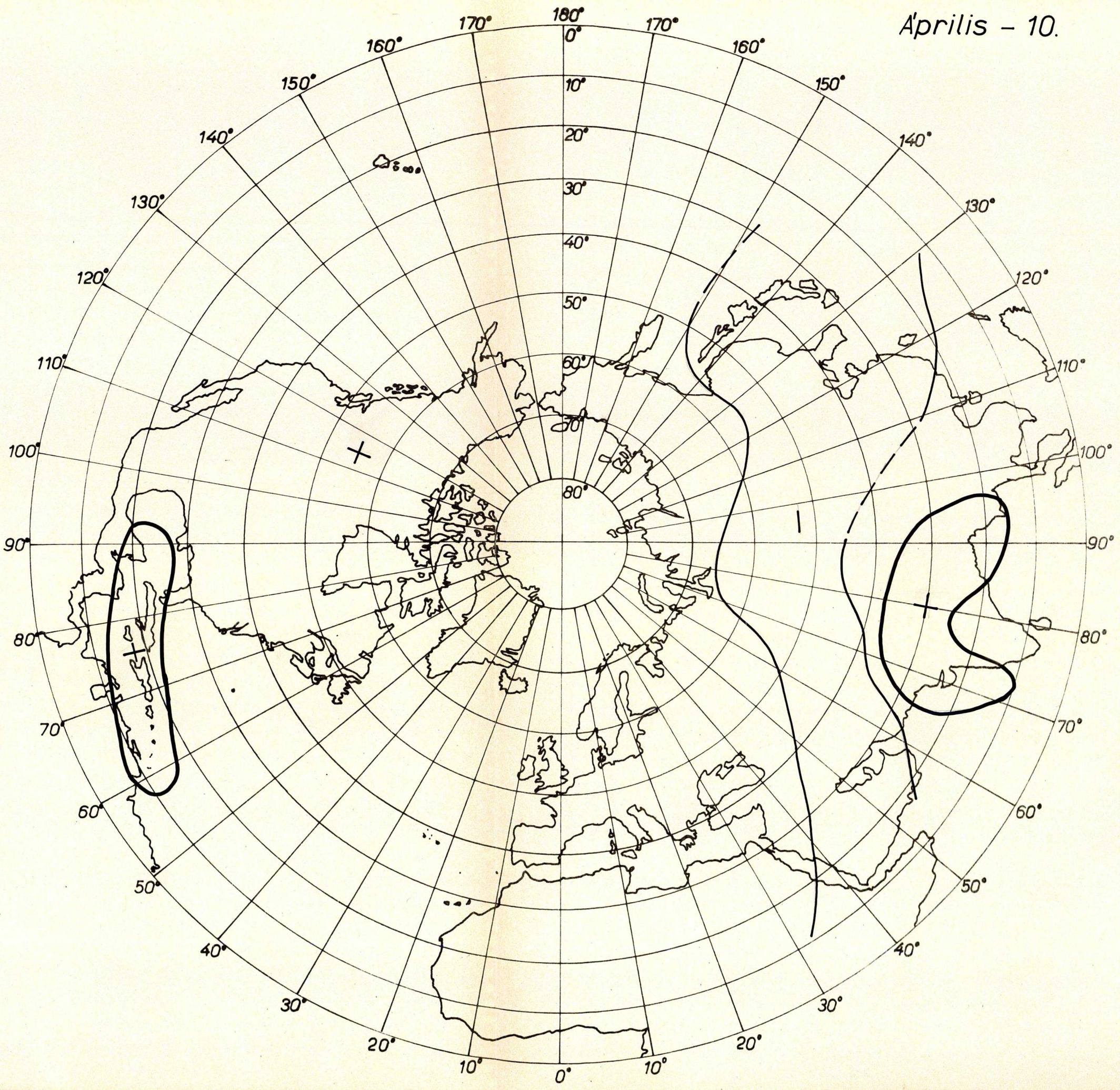
Aprilis - 2.



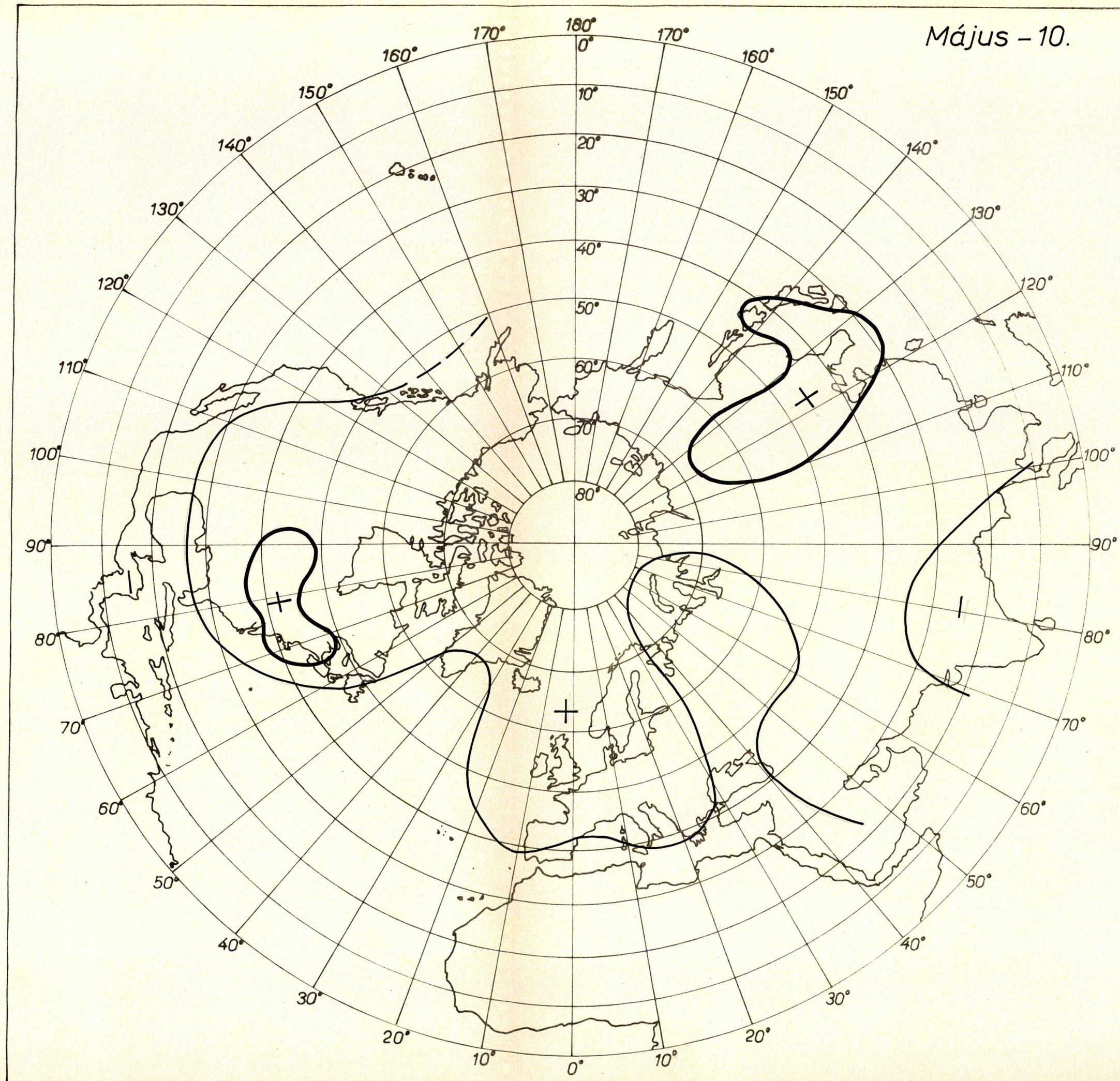
Aprilis - 7.



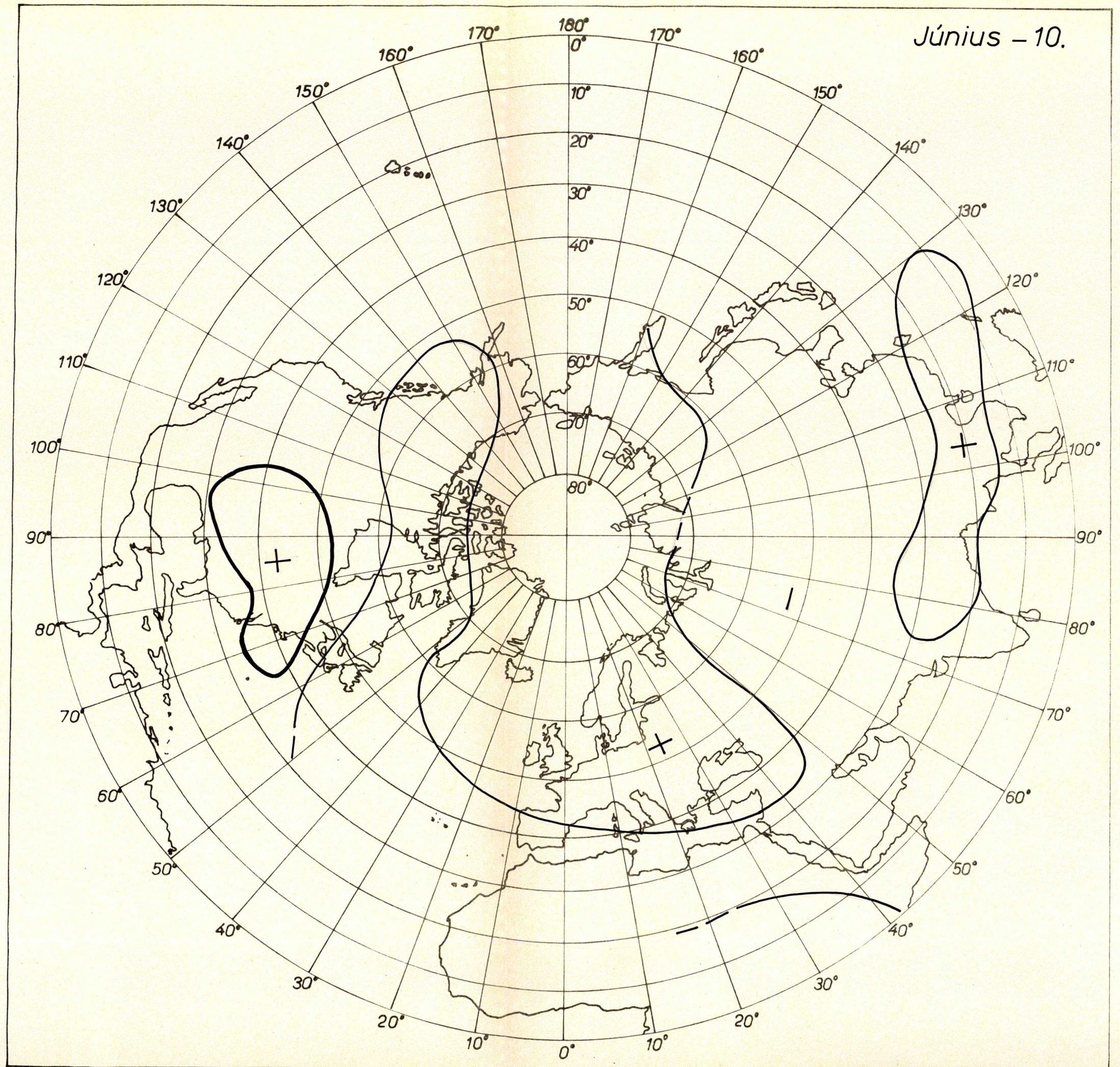
Aprilis - 10.



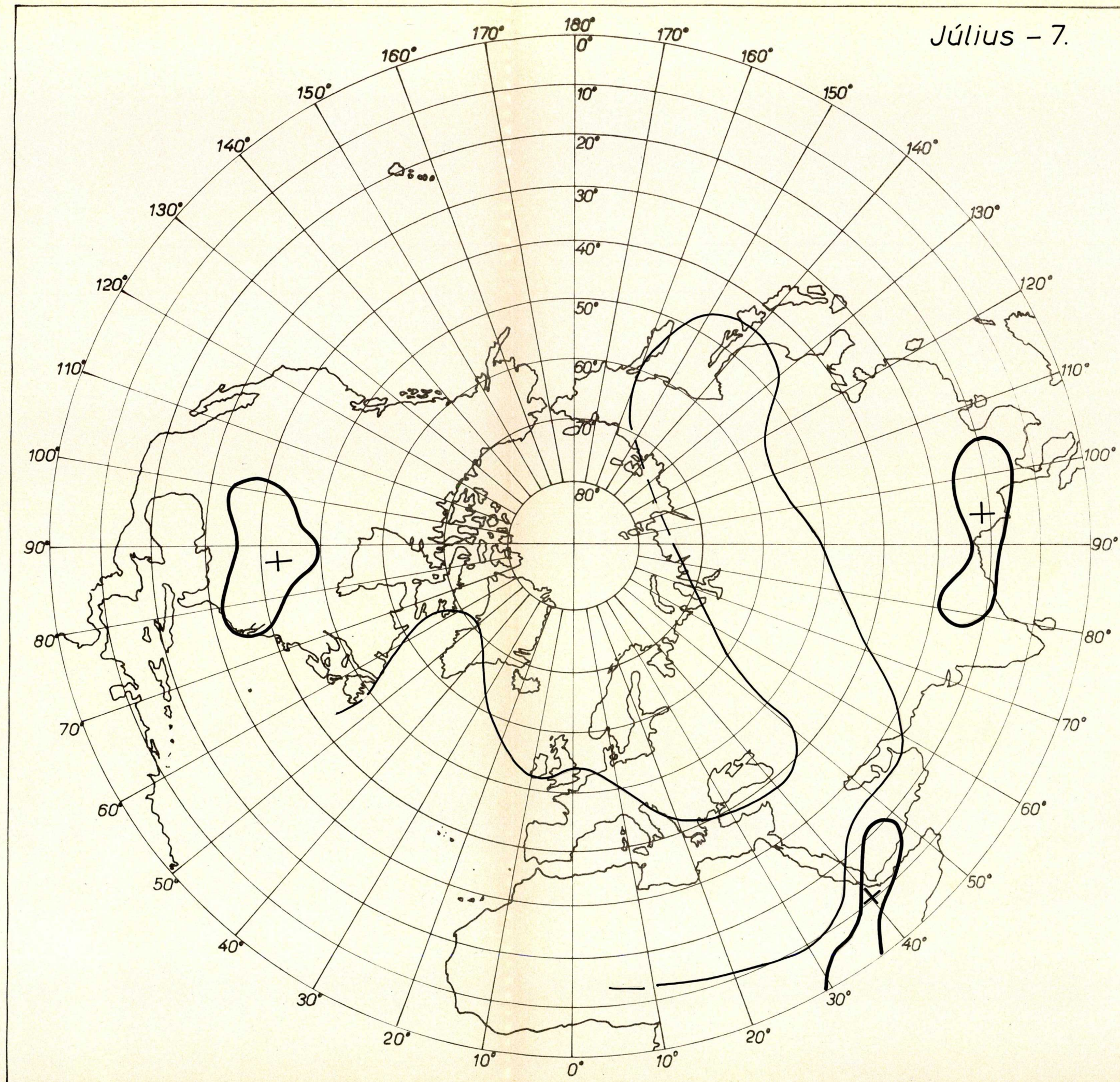
Május - 10.



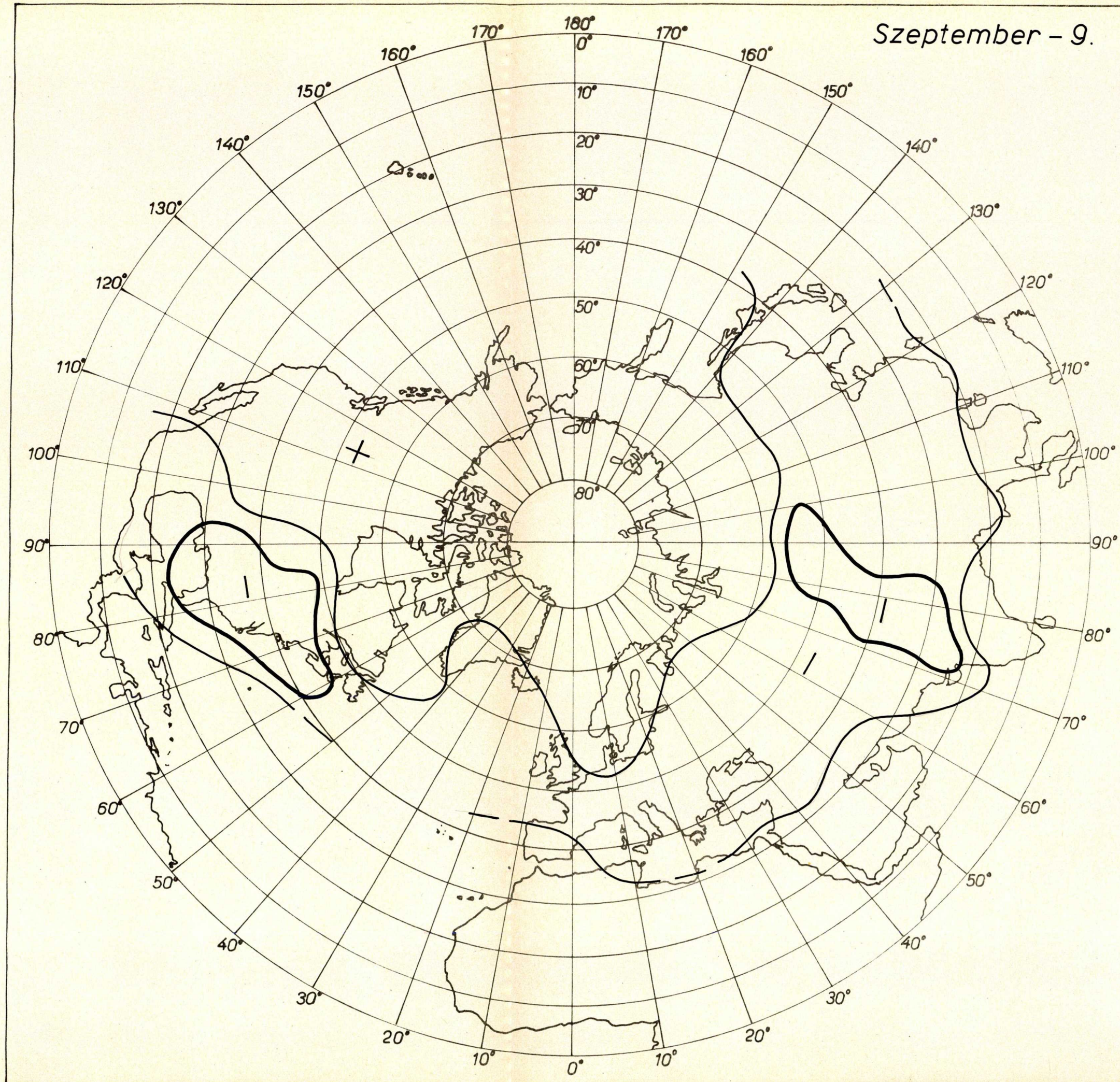
Június - 10.



Július - 7.



Szeptember - 9.



Október - 9.

