FOREWORD

Last winter, Dr. William Warntz, research associate with the American Geographical Society, presented a guest lecture to the members of MICMG. In that lecture, Dr. Warntz spelled out quite clearly, among other things, the necessity for a closer association by geographers with its sister geo-science, geometry, in order to more fully understand the complexities of spatial distributions.

Since Dr. Warntz's comments appeared particularly germane to the interests of MICMG discussion paper readers, he agreed to expand this topic into a formal discussion paper. Thus, MICMG is pleased to present, as our second guest author, Dr. Warntz's views on an essential and timely topic in the expanding area of mathematical geography.

The Editor, June, 1965

A NOTE ON SURFACES AND PATHS AND APPLICATIONS TO GEOGRAPHICAL PROBLEMS

By William Warntz

Recent gains in the development of general geography as a science at the theoretical - predictive level have been accomplished largely, perhaps, as a result of the simultaneous increase in both the sophistication and the naiveté with which geographers view the phenomena of the real world. This seeming paradox is explainable, for example, with reference to mathematics. More than ever before, geographers are using the tools of calculus, probability, topology, symbolic logic, the various algebras, geometries, for example, are being taken more literally than ever before.

We are aware of the topological nature of problems in non-spatial sciences like chemistry involving connectives or bondings of molecules, and of those concerning hierarchial chains of command, responsibility, and interactions as recognized in certain of the non-spatial social sciences. In addition, in general classificatory science and in the mathematical set theory underlying it, use is made of Venn diagrams, which utilize topological properties of an idealized space to portray graphically such relations as are implied in subsets, intersects, and unions, etc. But, we can take Venn diagrams in a far more literal sense than they were originally intended and by substituting real space and attendant phenomena for ideal space and by insisting upon utilization of all of the geometric properties involved as well as just the topological ones, geographers can reinterpret, add to, and refine the conventional concepts in the methodology of uniform regional geography and provide it with a basis in logic. Further discussion of this important topic lies outside the present paper and is reserved for presentation elsewhere. Of course, geographers also take the spatial considerations in topology literally in analysis of highway networks, river systems, and so on.

With regard to geometries, one also can cite numerous examples of geometrical solutions to problems which are not inherently spatial or in which the problems have been abstracted from space, and geometry is employed only by analogy. Included would be such things as one approach in economics to consumer tastes, prices, and preferences. Indifference curves are used to portray a surface of satisfaction. Various paths on this surface have meaning with regard to the income effect and the substitution effect. Other examples abound in economics.

In chemistry the use of surfaces and paths to show relationships in non-spatial thermodynamics is due to the nineteenth century American scientist, Josiah Willard Gibbs. His methods of geometrical representation of thermodynamic properties of substances by means of surfaces showed, for example, how to diagram water as it undergoes changes from solid to liquid to gas. So impressive was this work that the brilliant British scientist, Clerk Maxwell, saluted Gibbs by building for him a plaster model entitled, "a statue of water." In general the mathematics of response surfaces, supported by appropriate statistical measures now appear as well in many non-spatial sciences, e.g., psychology, learning theory, and so on.

Today geographers are taking the geo in geometry literally and the study of earth related surfaces and paths has now been expanded far beyond its original application to such things as land forms, contour mapping, drainage patterns, temperatures, pressures, precipitation, and the like in physical geography alone.

The modern geographer conceives of surfaces based also on social, economic, and cultural phenomena portraying not only conventional densities but other things such as field quantity potentials, probabilities, costs, times, and so on. Always, however, these conceptual surfaces may be regarded as overlying the surface of the real earth and the geometric and topological characteristics of these surfaces, as transformed, thus describe aspects of the geography of the real world.

Elaboration of additional important ideas may be found in William Bunge's Theoretical Geography, Lund, Sweden, 1962. Bunge has pointed out the necessity and the efficiency of recognizing the inseparability of geometry as the mathematics, i.e., the language, of spatial relations and geography as the science of spatial relations.

Illustrations of the application of the ideas, the geometry of surfaces and paths to a number of geographical problems can readily be given. The examples following have been selected because of their non-spatial diversity.

Our first example concerns any conformal map projection with the one exhibited here in figure 1, the well known Mercator projection - equatorial case. An exact mathematical isomorphism exists between the paths of light rays in an isotropic medium with an index of refraction varying from point to point but constant in all directions around any given point and the least distance paths or great circle arcs on the earth's spherical surface as represented on a conformal map. Let f be the scale of a conformal map at any point - expressed as the fractional ratio of distance on the map to actual distance on the earth, and let ds be the infinitesimal line element measured on the map. Then, the value $\int \frac{1}{f} ds$ is a minimum for the great circle track between any two given points as compared with the values obtained by integrating along any nearby alternative paths. This can be stated as a calculus of variations problem.

The designation of great circle tracks can also be accomplished by graphical portrayal of a surface and a gradient path. Shown in figure 1 is a Mercator map with a distance surface based on London depicted by isodistance lines in statute miles with a constant interval. On the real earth such lines would be concentric circles at first with increasing circumferences and then decreasing to zero at the antipodal point. When portrayed on the Mercator projection these circles are transformed, for although the Mercator is conformal in the small, it cannot show

DISTANCE SURFACE AND GEODESICS BASED ON LONDON

MERCATOR PROJECTION

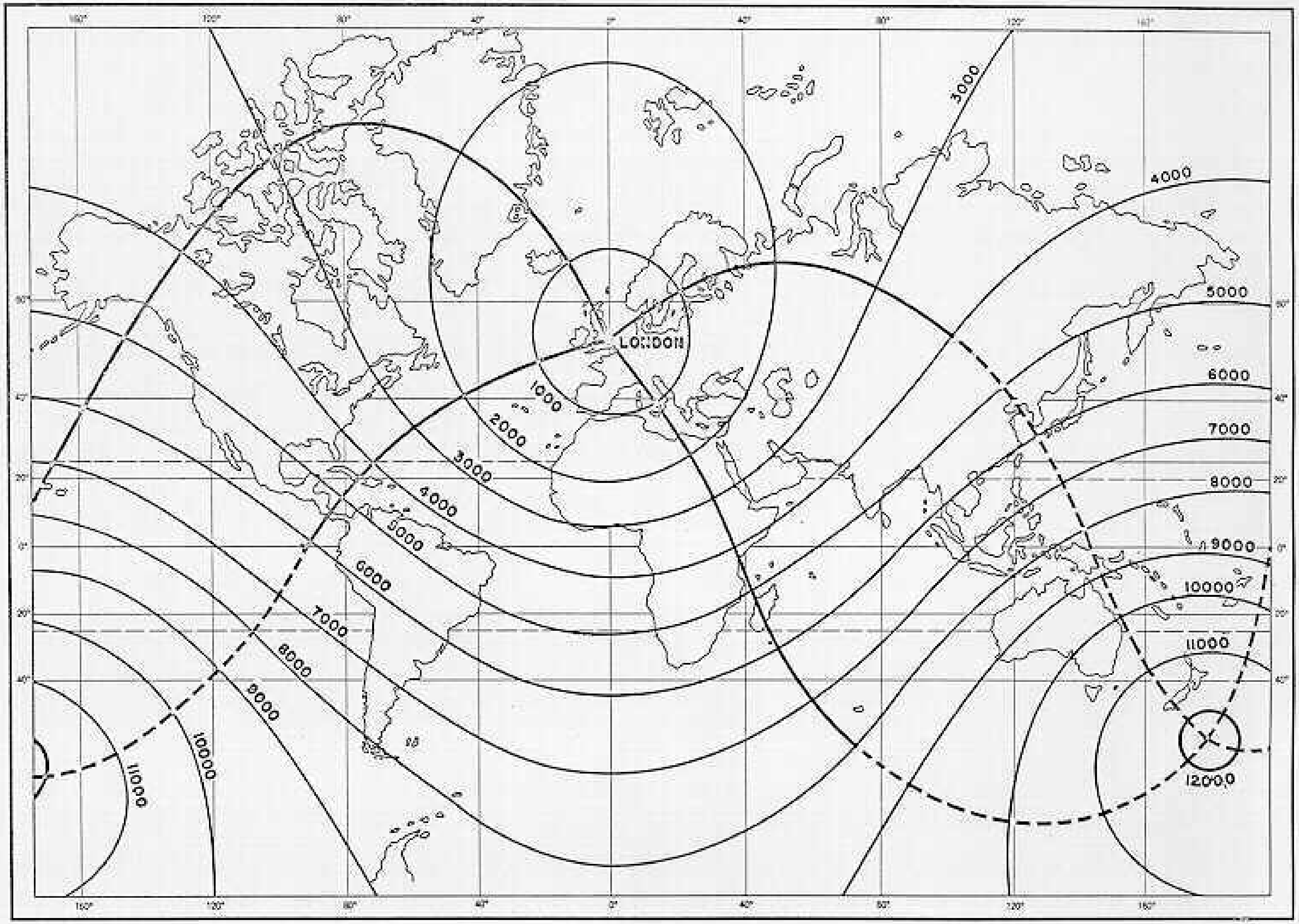


FIGURE I

correct shapes in the large because of scale change. On a spherical earth great corcles from all points to a given point would be orthogonal trajectories to the iso-distance circles centered on the given point. On this conformal map that property is retained but the great circles must bend to achieve it. Notice then, the presentation in figure 1 of the isolines of the distance surface and the orthogonal trajectories or gradient paths providing one solution to the problem of great circle determination. (Note that an actual geodesic on the map plane, the straight line, traces out a rhumb or constant heading line on the earth.)

Figure 1A shows another conformal map - a stereographic with the center of projection, in this case, at the South Pole. Shown on this map are certain iso-distance circles in statute miles and great circle paths based on Salisbury, Southern Rhodesia. Again the gradient path for great circles on the distance surface is found. The rule of orthogonal trajectory applies to this as to all other conformal maps.

The stereographic has a number of interesting properties. All great or small circles, or arcs of circles, on the earth's surface map as circles or arcs of circles on the stereographic. But concentric circles on the earth, while mapping as circles, do not, in general map as concentric circles. The exception, of course, is the case for the center of the projection. In figure 1A (assuming a perfectly spherical earth) small circles of latitude (these may be regarded as iso-distance lines on the earth) do map as concentric circles about the South Pole. However, let us look at the iso-distance circles about Salisbury. These are concentric circles on a spherical earth but on the plane stereographic projection they map as non-concentric circles - or non-concentric arcs of circles since the entire earth is not shown on one stereographic projection.

The centers of these mapped iso-distance circles about Salisbury march steadily away from Salisbury along the straight line on the plane of the projection from the

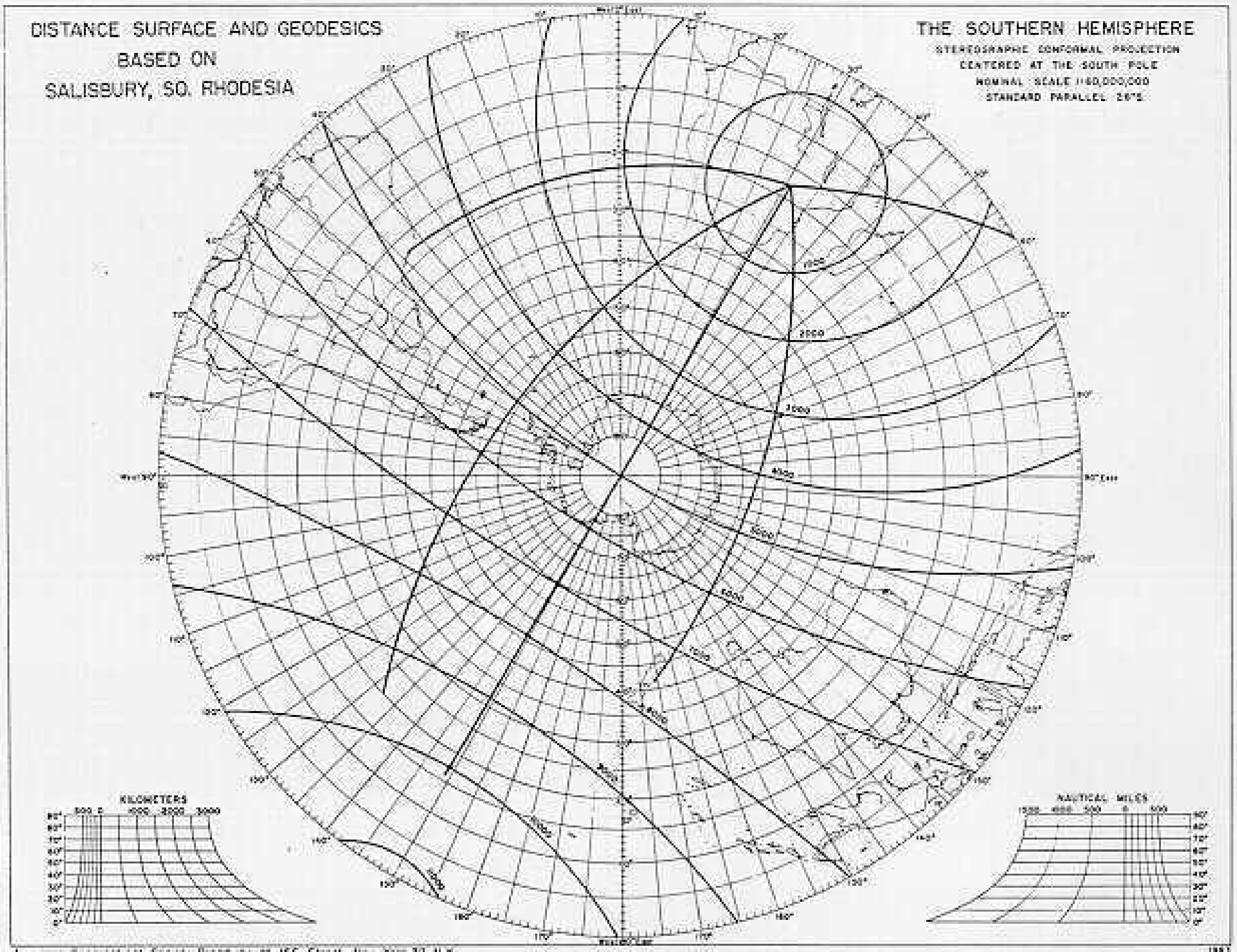

DISTANCE SURFACE AND GEODESICS
BASED ON
SALISBURY, SO. RHODESIA
THE SOUTHERN HEMISPHERE
STEREOGRAPHIC CONFORMAL PROJECTION
CENTERED AT THE SOUTH POLE
NOMINAL SCALE 1:60,000,000
KILOMETERS
NAUTICAL MILES
American Geographical Society, Broadway at 156 Street, New York 32, N.Y.

FIGURE IA

center of the projection (the South Pole) extended through Salisbury. The direction of movement, of course, is away from the center of the projection. The iso-distance circle that passes through the antipodal point of the center of the projection will map as a straight line. For this map the antipodal point of the center of the projection is the North Pole. (This is also the center of perspective for this projection.) So, the iso-distance circle about Salisbury that passes through the North Pole maps as a straight line and its center on the plane of projection lies at infinity, in fact at both plus and minus infinity if we adopt that convention. The next iso-distance circle out from Salisbury will map with negative curvature and its center on the plane of the projection will lie between the point at negative infinity and the antipodal point of Salisbury. The centers of the succeeding iso-distance circles based on Salisbury will march ever closer to Salisbury's antipodal point. This path, of course, is the straight line on the plane from the point of minus infinity to the antipodal point of Salisbury and which when extended would also pass through the South Pole (and beyond through Salisbury and to the point of plus infinity noted above).

Once the positions of the selected iso-distance circles have been determined the drawing of the least earth distance path or geodesic from any point to Salisbury is a simple matter. It is the orthogonal trajectory or gradient path on this conformal map. All places are relative sources for Salisbury as the single sink. The refraction analogy holds again and the reciprocal of the map scale at any point may be regarded as an index of refraction for the geodesic there.

It is interesting to note that the family of geodesics passing through Salisbury which, of course, are great circles on the earth's surface also map as circles on the stereographic.

On the surface of the spherical earth, the latitude and longitude circles may be regarded, respectively, as distance circles about, and geodesics passing through, the geographic poles. Thus, our Salisbury case resembles the polar case on any non-polar stereographic.

The next example considers the problem of determining the path for minimum time enroute for an aircraft between two airports separated by a broad expanse with winds of varying direction and velocity. Almost never will the least distance path, i.e., great circle route, afford the minimum time enroute. Generally, it is possible to deviate from this great circle route to enhance the speed over the earth's surface by utilizing more favorable winds. This will be attempted, ordinarily, so long as the addition to speed is proportionally greater than the addition to distance. An elegant mathematical solution to this calculus of variations problem exists which, however, is not practical. Essentially the problem is one of the path which minimizes the value, $\int \frac{1}{v} ds$ when v is the speed of the aircraft over the earth's surface and ds is as above. Difficulty arises from the fact that velocity and direction of the aircraft over the surface depends upon the vector of heading and true air speed of the aircraft combined with that of wind direction and velocity. The resulting tail wind or head wind component at any given point in the wind field is not independent of the aircraft's actual heading. But, a graphical solution can be used to develop a family of isotime lines and one also of course lines based on a given departure point and showing the times and routes for minimum time flights to all destinations. The method involved is that modified by including wind vector considerations from the seventeenth century Dutch scientist, Huygens, relating to sound and light refraction based on wavelets, envelope curves, and time fronts. Figure 2 shows some of the family of minimum time routes on the time surface integrated about New York City for DC-8 jet aircraft flying to Europe at a constant pressure altitude of 300 millibars on October 17, 1960, and Figure 2A shows how to create such a surface and attendant paths.

When portrayed in full, the situation depicted in figure 2 reveals cusps or

TIME SURFACES AND ROUTES FROM NEW YORK CITY, DC-8, OCTOBER 17, 1960

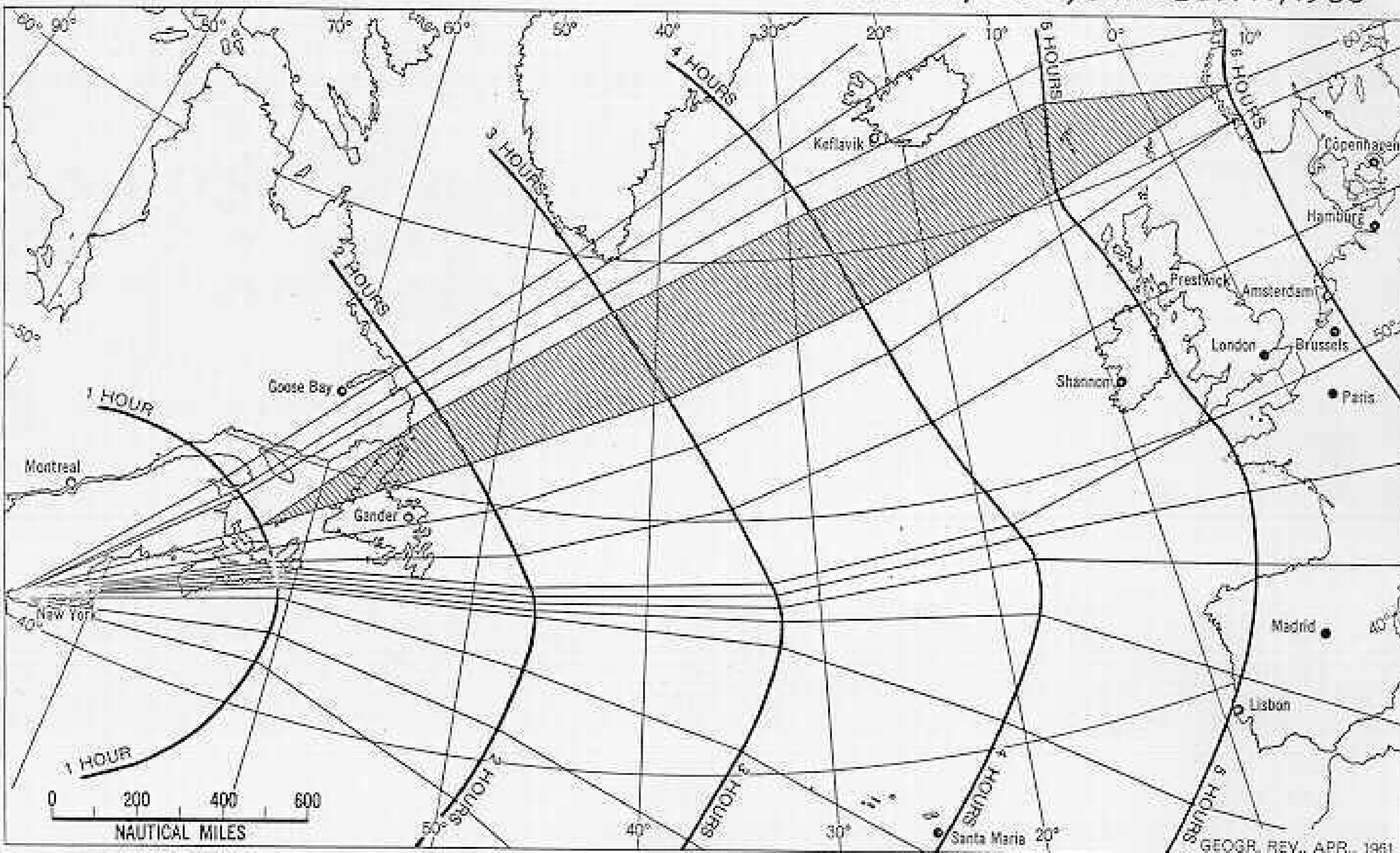
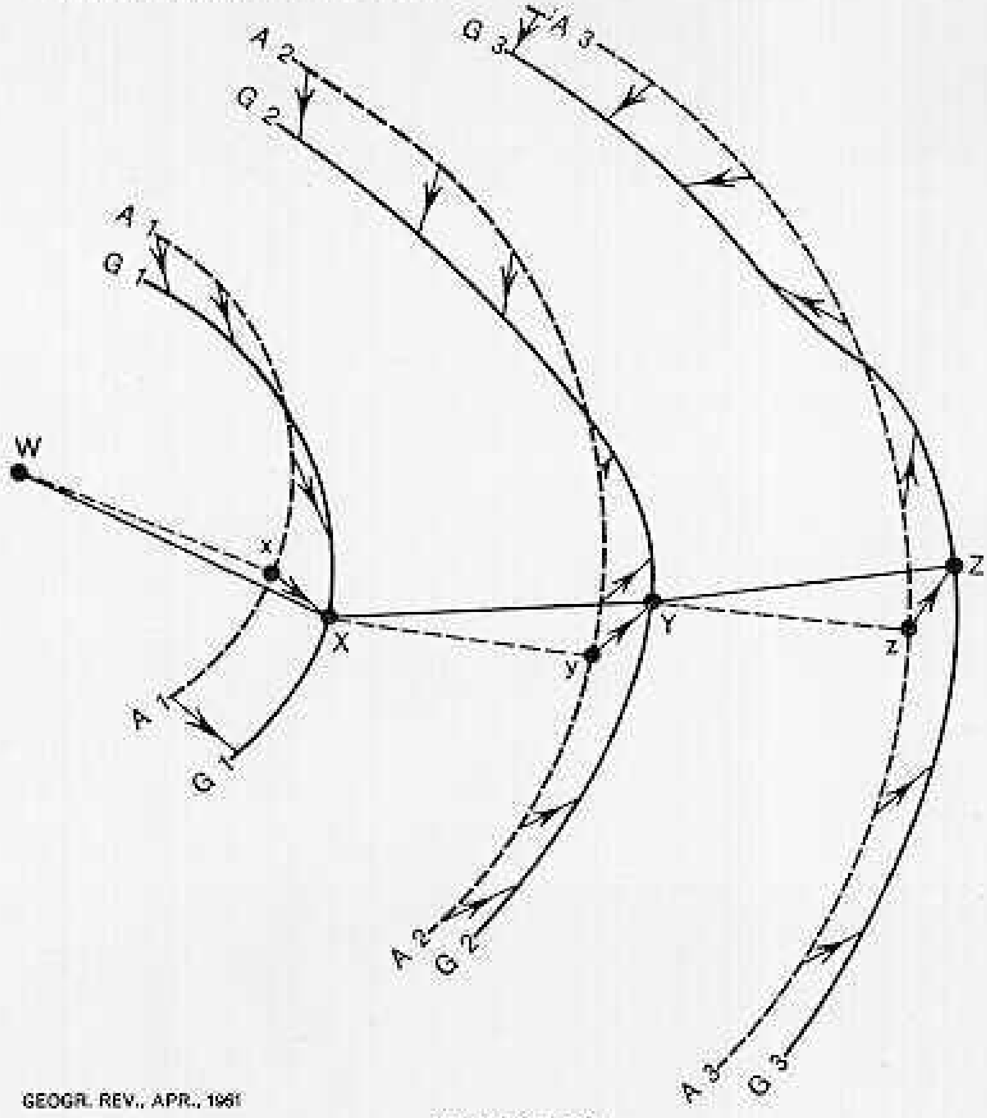


FIGURE 2

GRAPHICAL DETERMINATION OF LEAST TIME PATH



GEOGR. REV., APR., 1961

FIGURE 2A

focal points in the least time routes and thus provides alternate equal minimum-time paths to certain destinations. Such a situation refers to the induction of caustics and occurs at the rear of a closed vortex associated with a strong high or low or in a region of strong wind shear as in a jet stream or a frontal zone.

Under certain pressure distributions complicated systems of time fronts and routes exist, and, as with their isomorphic counterparts in optics, reflection, diffraction, refraction, and other patterns can be ascertained to exist in addition to, and in agreement with, the caustics.

With departure point W in figure 2A as the center, draw a circle with a radius representing the distance the aircraft can fly in one hour at its cruising true air speed. Only an arc of this circle is shown here, covering the general direction of the intended flight. This arc of the circle, labeled A1, indicated the maximum distances the aircraft could fly in the complete absence of wind. The effect of wind is estimated by drawing the appropriate wind vectors representing velocity and direction from a number of points on the air position line A1. The heads of these wind vectors can be connected by the smooth curve G1, which designates the position of the time front after one hour and gives the farthest ground positions the aircraft could achieve in that time in the existing wind field.

From as many points as desired on curve G1, arcs may be swung off each with a radius again equal to the true air speed. The curve A2 is drawn as the envelope of these arcs. Displacement of A2 by the appropriate wind vectors yields G2, the position line of the time front at the end of the second hour after departure.

This procedure can be continued until the time front closest to required destination Z is obtained. In the case portrayed here destination lies exactly on a time front. If, as is usually the case, destination does not lie on a time front, interpolation will yield an adequate estimate of the least time in which the flights can be accomplished.

To obtain directly the specific route form W to Z that affords this minimum flight time, plot backward from destination to departure point. From Z (in this case on G3) the wind vector is plotted in reverse, giving the point z on A3. From this point the perpendicular is dropped to G2. This new point, Y, is on the required optimum flight path.

Repetition of this procedure yields point X on G1 and finally W as the departure point. The optimum flight path, here the minimum-time path, is then the curve WXYZ. This course represents the connected series of ground-speed-true-course vectors each perpendicular to its attendant air position curve. Likewise, the true-air-speed-true-heading vectors are perpendicular to their ground position curves, the time fronts.

The methods used here result in approximations of the desired result. Such graphical procedures involve arbitrary discontinuities and averages. Specifically, the accuracy achieved apart from the adequacy of the meteorological forecast tends to vary inversely with the length of the time interval chosen to portray wind velocities, air speeds, ground speeds, and the successive positions of the time front.

The final examples utilize a potential of population map based on 1960 census counts for the contermininous United States and computed and compiled jointly by the American Geographical Society and Princeton University using the IBM 7090 computer. Figure 3 shows this map, the most detailed such map yet produced. Potential of population is shown as a macrogeographic spatially continuous surface by means of isolines indicating geographical variation in the values on this surface. Potential of population measures the aggregate accessibility of the entire population of the country to all points when the value at any given point represents the summed contributions from all members of the population when each contribution is proportional to the reciprocal of the distance of the person away from the given point. More specifically, the value of potential of population at a given point is $\int \frac{1}{r} D\,dA$ when D is the population density over any infinitesimal element of area, dA, and r is

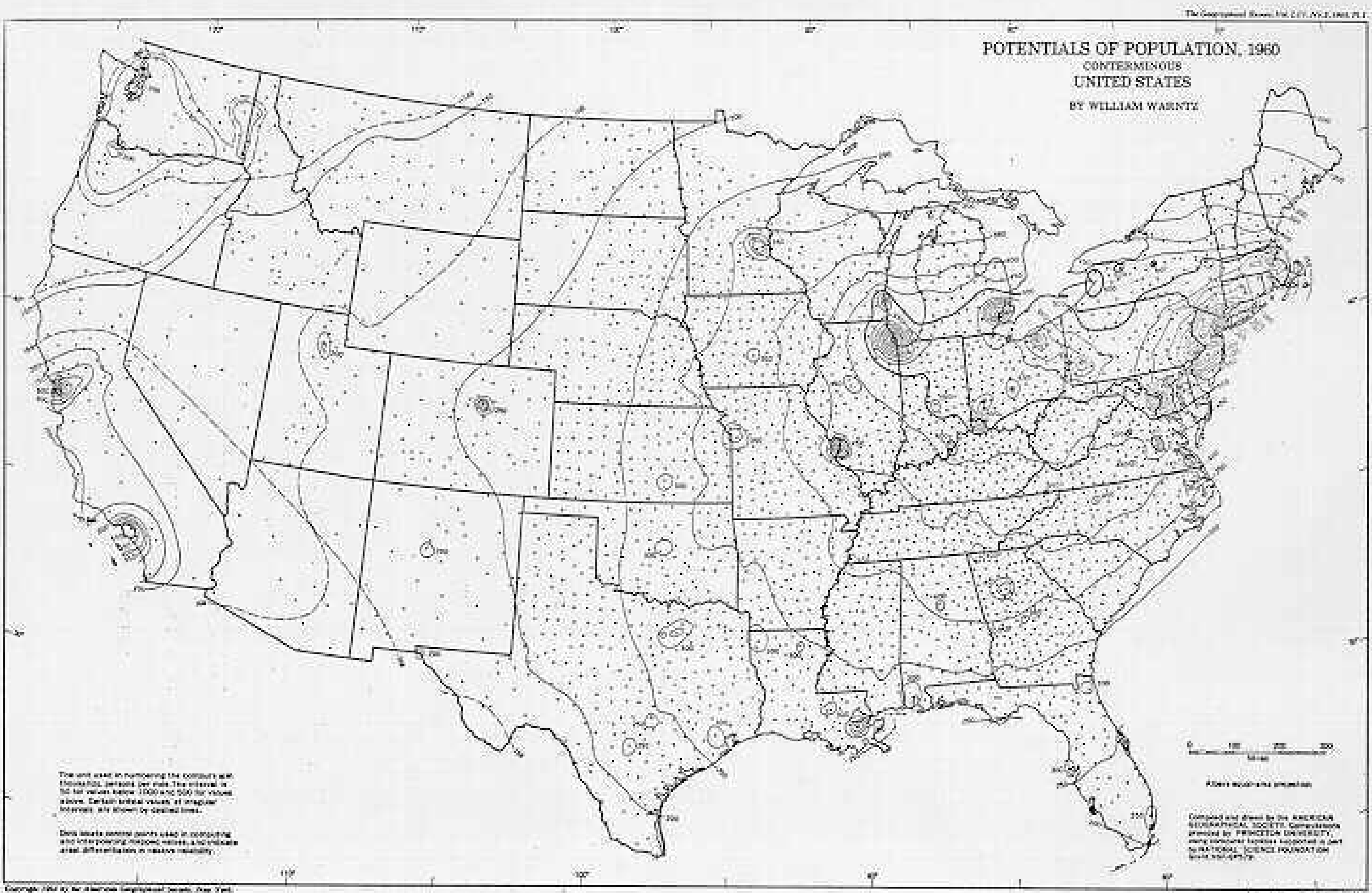

POTENTIALS OF POPULATION, 1960
CONTERMINOUS
UNITED STATES
BY WILLIAM WARNTZ

FIGURE 3

the distance of each such element from the given point. The integration is extended to all elements when D is not zero. The value of potential of population can be determined for as many points as required to facilitate mapping the surface by the contouring technique. Units of potential of population are in persons times distance to the minus one power. On figure 3 values are in thousands, persons per mile. To convert to thousands, persons per kilometer, the mapped values may be multiplied by the factor 0.622.

Maps of this sort indicate spatial structuring of a wide range of economic and social phenomena and are based upon formalization arising from empirical formulas describing certain spatial processes, i.e. flows that are of economic and social importance.

With regard to spatial structure it is important to note that values of non-urban land in the United States vary directly with and in close agreement with the potential of population surface. Specifically for 1960, land value in dollars per acre = 6.104^{-6} times potential of population raised to the 1.6 power when potential was in units of thousands, persons per mile. The coefficient of correlation was 0.863 based on state averages and thus was quite high. For the number of degrees of freedom obtaining here any value of the coefficient of correlation exceeding 0.288 is to be deemed significant at the five per cent fiducial level.

Urban land values increase with potential to an even higher power than for non-urban land so that one simple linear logarithmic function is inadequate. Even so, however, one may closely approximate a continuous land value surface for the United States by transforming the potential surface everywhere by the factor of proportionality and exponent of potential given above due to the peaking of potential locally in urban areas. Such a procedure will permit us to provide the illustration intended below.

Let us imagine that it is the task to establish from any given place the routes to all other places in the country and in each case the route is the one for which the total land acquisition cost would be at a minimum. A map of such a family of routes and the attendant isocost surface on which these routes represent gradient paths can be produced by means of the simple Huygens' graphical method. Figure 4 represents an isocost surface integrated about Lewistown, Montana. Lewistown, in this case, is a "sink" and all other places are relative "sources." The minimum land acquisition cost route to any other given point in the United States may be found by plotting the orthogonal trajectory to the isocost lines from the given point back to Lewistown. Hence, the routes are gradient paths. Several such paths are shown on the map. Note the divide between the northerly and southerly routes emanating from this town in Montana. To further emphasize the nature of such surfaces and the paths on them, an additional graphical presentation is supplied in figure 5 based on Murfreesboro, Tennessee.

Note particularly by comparing figures 4 and 5 that the path from Lewistown to Murfreesboro is precisely the same as that from Murfreesboro to Lewistown. Thus, although the sets of iso-cost contours differ greatly the tangents to both sets of contours coincide along this path. Because the original values on the land value surface are independent of the directions of lines through points on that surface, the conditions obtains that between any two points the least cost paths coincide. Other than this, however, a separate map is required for each point to obtain the minimum cost paths from all other points to the given point. The possibility of producing one general map to accomplish all such paths is being investigated.

For the minimum time paths for aircraft as shown in figure 2 the same path does not exist in opposite directions between any two points. As noted earlier, the effect at a point of the wind on the speed and the heading of an aircraft depends upon the heading and the speed with which the aircraft approaches the point.

COST SURFACE AND MINIMUM LAND ACQUISITION COST ROUTES BASED ON LEWISTOWN, MONTANA

(IN HUNDREDS OF THOUSANDS OF DOLLARS)

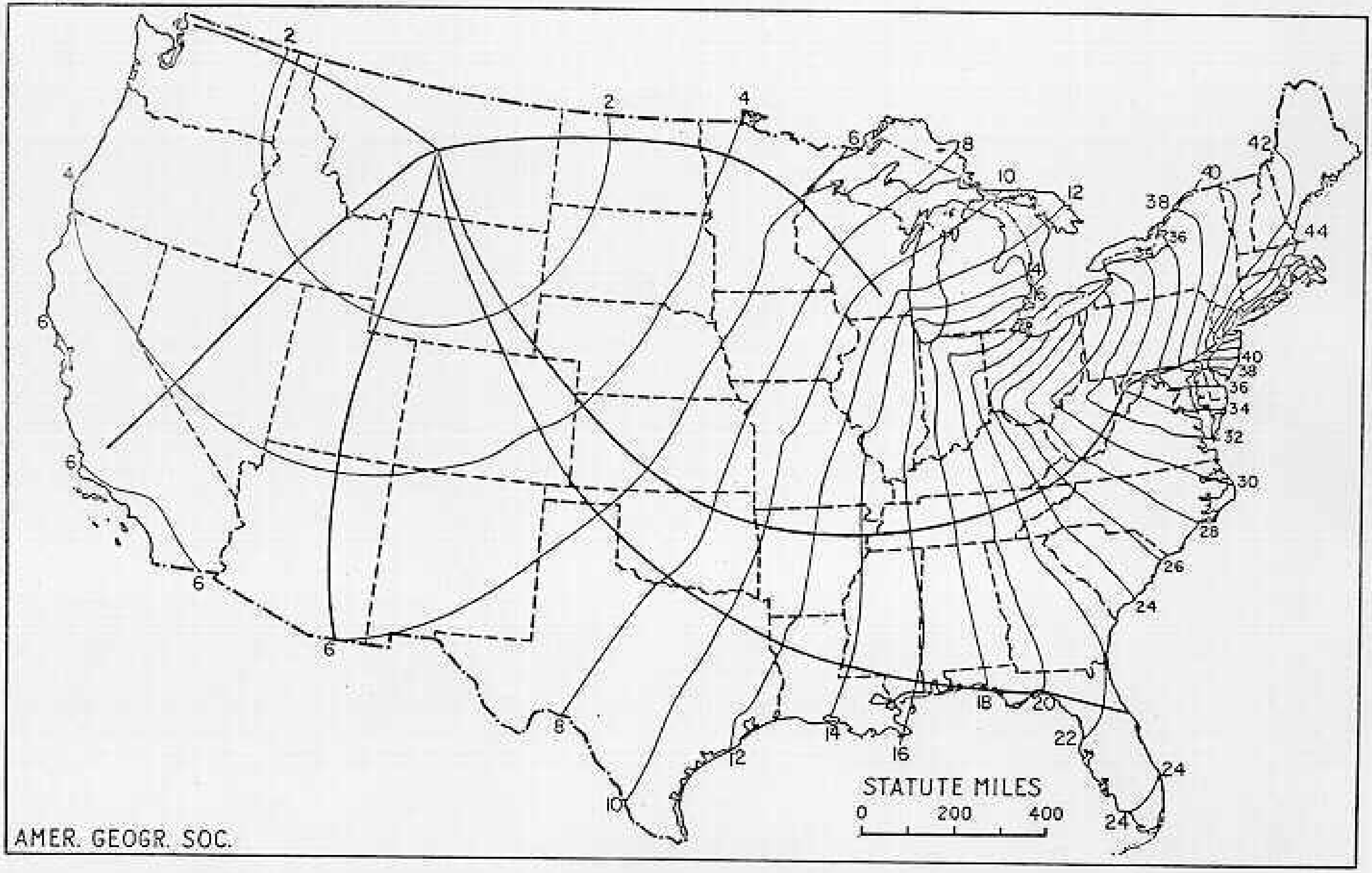


FIGURE 4

COST SURFACE AND MINIMUM LAND ACQUISITION COST ROUTES BASED ON MURFREESBORO, TENNESSEE

(IN HUNDREDS OF THOUSANDS OF DOLLARS)

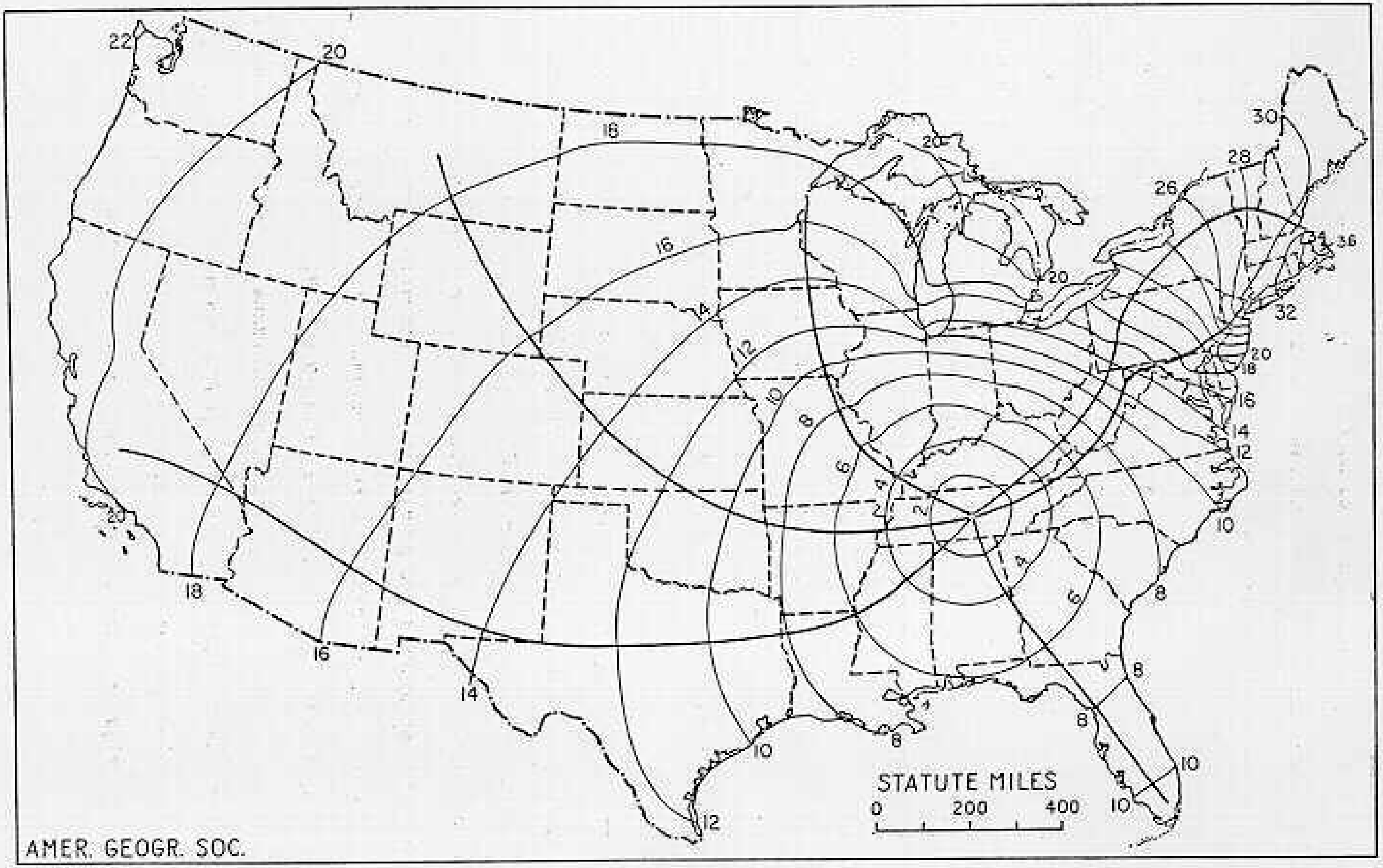


FIGURE 5

The procedure used to achieve the surfaces represented in figures 4 and 5 involved a decision as to the finite width assumed for the minimum cost routes to be shown. This is indicated in the legend of figure 6 showing the graphical method employed to determine these surfaces and paths.

Many other examples might be given in addition to the three kinds given above, but the distance surfaces, time surfaces, and cost surfaces that these represent are indicative of the wide application of the basic ideas of surfaces and paths in geography generally. This, of course, includes a part of the general theory of geodesics, and intensive study and the literal application of it to the real earth and especially the various conceptual surfaces covering it seems called for at this stage in the development of geography and its growth toward a truly unified discipline.

We have noted that mathematical and graphical solutions exist for our kinds of problems. In some cases, the mathematics is intractable, however, and graphical solutions, or rather approximations, do not exist or have not yet been invented.

A third kind of solution exists as an extension of the graphical method and that is to build three dimensional models of the various surfaces. For example, one could make a model of the equatorial mercator projection of the iso-distance surface based on London. A convenient scale could be established between values on the distance surface and height above the plane on which London was regarded to exist. Thus London would become the pit with all others points elevated above it and with heights increasing in a monotonic and continuous fashion outwards until the one peak on the surface, the antipodal point for London was reached.

One could make this into an operational model, indeed a geographical analogue computer, for finding the great circle path on the surface from any point through London by allowing a ball to roll freely on the surface from that point. It will roll "down hill" to the pit of London along the gradient path. A difficulty exists

GRAPHICAL METHOD FOR DETERMINING COST SURFACE ABOUT A POINT

S AND S' REPRESENT SUCCESSIVE ISOCOST LINES; $r_1 r_2 r_3$ = DISTANCE RADII ACHIEVABLE FOR OUTLAY OF $100,000 ASSUMING THAT 15 ACRES OF LAND ARE REQUIRED FOR EACH MILE OF HIGHWAY LENGTH

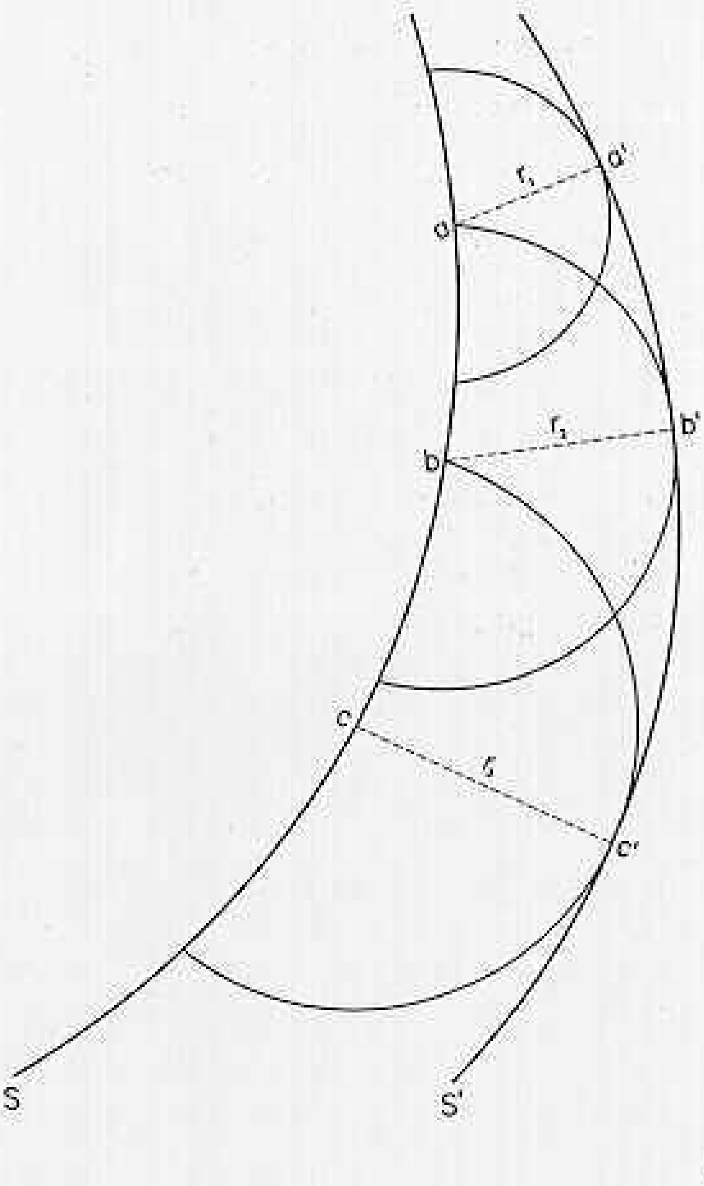


FIGURE 6

because the ball has mass and gained momentum may cause it to depart from the locally steepest slope at some places. However, if the model's vertical scale is selected so that no slopes are too steep the problem is minimized. Introducing additional friction to the surface could help also.

Boundary problems exist also. The interruption of the map along a meridian is easily overcome by making the model periodic in an east-west direction. The fact that the geographical poles lie at infinity poses another less easily solved problem.

Models for the other phenomena presented in this paper are possible too. Thus, elevation on the surface could be made proportional to time from New York, or cost from Lewistown, etc. Our geodesic-finding ball would approximate the required paths.

Geography, geometry, and graphics which had their first grand synthesis in cartography at the time of Ptolemy stand to benefit mutually in the recently established cooperation which greatly extends and intensifies those early benefits. This time spatial patterns in general are being considered with regard to social, economic, and cultural phenomena as well as physical phenomena with recognition that the geo in geometry and the geo in geography have more in common than we would have dared to dream a decade ago.

REFERENCES AND ACKNOWLEDGEMENTS

Part of this paper was presented to the International Geographical Congress, London, July 1964. The topics in it were also presented to the faculty and students of the Michigan Inter-University Community of Mathematical Geographers in November 1964. The author is deeply indebted to that group for their incisive comments and recommendations.

All figures have been drawn by the cartographic staff of the American Geographical Society. Figures 1, 1A, 4, and 5 appear here for the first time. The others have appeared previously in the *Geographical Review* and are reproduced here with permission.

Thanks are owed to the Editor of the *Geographical Review* for allowing the author to quote passages from his article, "Transatlantic Flights and Pressure Patterns," from Vol. 51, No. 2, 1961, pp. 187-212 of that journal.

John Z. Stewart in "The Use and Abuse of Map Projections," *Geographical Review*, Vol. 33, No. 4, 1943, pp. 589-604, pointed out the analogy between the inverse scale on any conformal map and the index of refraction in geometrical optics.

Various staff members at the American Geographical Society have advised the author. In particular, O. M. Miller, the assistant director of the Society, offered valuable advice.

At Princeton University Professor Steve Sláby and Mr. C. Ernesto Lindgren of the School of Engineering's Department of Graphics and Engineering Drawing have provided many enlightening comments. They are much interested in the modern revival of the cooperative assistance of geometry and geography and especially of the role that graphics plays in this.

Also at Princeton University, the sculptor in residence, Joe Brown,

advised the author and demonstrated for him certain of the techniques of three-dimensional model building. It was in Mr. Brown's studio and with his advice that the author completed the three-dimensional model of the 1960 Potential of Population Surface for the United States. This model is currently included in the American Geographical Society's exhibit at the New York World's Fair.

The gentlemen cited above must in no way be held accountable for the defects of this paper, however. Those came from the mind of the author alone. But, the author is deeply indebted to these many gentlemen and would like to make known his gratitude to them.