

Sasaki's Pivotal Contribution: Calculus of Variations Applied to Weather Map Analysis

JOHN LEWIS

National Severe Storms Laboratory, Norman, Oklahoma, and Desert Research Institute, Reno, Nevada

S. LAKSHMIVARAHAN

School of Computer Science, University of Oklahoma, Norman, Oklahoma

(Manuscript received 2 October 2007, in final form 10 January 2008)

ABSTRACT

Yoshikazu Sasaki developed a variational method of data assimilation, a cornerstone of modern-day analysis and prediction in meteorology. Fundamentally, he formulated data assimilation as a constrained minimization problem with equality constraints. The generation of this idea is tracked by analyzing his education and research at the University of Tokyo in the immediate post–World War II (WWII) period. Despite austere circumstances—including limited financial support for education, poor living conditions, and a lack of educational resources—Sasaki was highly motivated and overcame these obstacles on his path to developing this innovative method of weather map analysis. The stages of his intellectual development are followed where information comes from access to his early publications, oral histories, and letters of reminiscence.

It has been argued that Sasaki's unique contribution to meteorological data assimilation stems from his deterministic view of the problem—a view founded on the principles of variational mechanics. Sasaki's approach to the problem is compared and contrasted with the stochastic view that was pioneered by Arnt Eliassen. Both of these optimal approaches are viewed in the context of the pragmatic–operational objective analysis schemes that were developed in the 1950s–1960s. Finally, current-day methods [e.g., three- and four-dimensional variational data assimilation (3DVAR and 4DVAR)] are linked to the optimal methods of Eliassen and Sasaki.

1. Introduction

During the early to mid-1950s when numerical weather prediction (NWP) set the meteorological world abuzz, there came the attendant need to provide initial conditions for deterministic models—first the quasigeostrophic (filtered) models and later the more general and physically more meaningful primitive equation models. These methods for finding the initial conditions were labeled “objective” or “numerical weather map” analysis schemes in opposition to manual or “subjective” analysis methods that were the standard in the precomputer age. In the late 1970s, the terminology for weather map analysis changed again—from objective

analysis to data assimilation or more specifically four-dimensional data assimilation (4DDA), that is, analysis in space and time.¹

Amid the development of the pragmatic and computationally efficient objective analysis schemes, there came a less pragmatic and computationally more demanding methodology based on the calculus of variations—the calculus of variational mechanics. Although this methodology rested on a solid dynamical framework, it was not a viable option in the early days of NWP. Implementation necessarily waited for advances in computational power and improved optimization al-

Corresponding author address: John M. Lewis, Desert Research Institute, 2215 Raggio Pkwy., Reno, NV 89512-1095.
E-mail: jlewis@dri.edu

¹ The First GARP (Global Atmospheric Research Program) Global Experiment (FGGE) took place in 1978–79. This program was aimed at extending the period of useful numerical prediction beyond several days and was the impetus for 4DDA (see Smagorinsky 1978, 18–19).

gorithms. Yoshikazu (“Yoshi”) Sasaki² laid the theoretical foundation for this innovative approach to weather map analysis in the austere conditions of post-World War II (WWII) Japan.

In addition to discussing Sasaki’s method of data assimilation, we pay homage to the other contributions of comparable importance that have had bearing on the current practice of data assimilation at operational weather prediction centers worldwide. These discussions, including a comparison and contrast of the various strategies, follow the historical study of Sasaki’s pathway to variational analysis (sections 2–5).

2. View from afar

In April 1947, Sasaki enrolled as an undergraduate in geophysics at the University of Tokyo. Among every group of 1000 students who passed through the rigorous education in Japan,³ only 1 gained entry to the University of Tokyo or *saiko gakufu*, the supreme institute of learning (Schoppa 1991).

At this time, less than two years after the emperor announced unconditional surrender to the Allies (15 August 1945), the country lay in ruin and the American occupation had commenced. The occupation would last for nearly seven years. Conditions for the students were not unlike those for the general population—an absence of adequate housing and a limited supply of food. As recalled by Yoshimitsu Ogura, a fellow student at the University of Tokyo, who later gained acclaim for his work in turbulence and cumulus convection:

First of all, I could not find a place to live, since so many houses were burned down. I brought my futon to the lab, and cleaned up the top of a big table every night to spread my futon and slept there. There were half a dozen homeless fellow young scientists in the Geophysical Institute. And I was hungry all the time. With an empty stomach, you cannot think of science well.

(Y. Ogura 1992, personal communication)

Since geophysics was a subspecialty within the physics department, the typical geophysics student followed a course of study identical to that of the physics student. During the final year, the geophysics student was given courses in seismology, geomagnetism, oceanography,

and meteorology (usually one lecture hour per week). Sasaki took these courses in academic year 1949–50. His B.S. degree was awarded in March 1950 after three years of study.

In his final undergraduate year, Sasaki took a course in group theory, that important mathematical subject that is pervasive throughout physics, chemistry, and applied mathematics. Kunihiro Kodaira, a young assistant professor taught the course. As recalled by Yoshi:

Professor Kodaira often came to class late and sometimes the students would leave [before he arrived],⁴ you know. It was kind of strange to have so few students. But it was just amazing how great a mathematician he was. I especially remember the first day of class when he came in late, bowed, and then went to the blackboard and filled it with equations. It was amazing.

(Y. Sasaki 1990, personal communication)

Several years after Sasaki took this course, Kodaira received the Fields Medal.⁵

After obtaining his bachelor’s of science (B.S.) degree, Sasaki joined the group of graduate students at the university’s Geophysical Institute. The university provided little or no financial support for the graduate student. Thus, Sasaki and fellow students at the Institute earned money in several different ways: Sasaki tutored high school students as they prepared for the demanding university entrance exams, Kikuro Miyakoda tried to make money by dabbling in the stock market while teaching physics and earth sciences part time at two high schools, and Akio Arakawa and Katsuyuki Ooyama served as upper-air observers aboard two Japan Meteorological Agency (JMA) ships off the east coast of Japan. The Japanese government could not cut funding for these two destroyer-sized ships, the *X-RAY* and *TANGO*, since the observations were needed by the occupational forces (K. Ooyama 1992, personal communication). These jobs offered steady employment, but the work assignments were demanding, especially taking place in the wintertime when rough seas were the rule.

In the immediate postwar period, the meteorological community outside of Japan was making notable advances. Upper-air observations were being routinely collected over Europe and the United States, and this

² Sasaki was granted U.S. citizenship in 1973. At that time, he changed the structure of his name to read Yoshi Kazu Sasaki, appearing as Yoshi K. Sasaki in publications after naturalization.

³ This system had equivalency to the *gymnasium* in Germany or the *lycée* in France, but is now called the high school under the old system or *Kyusei Koko* (T. Tsukahara 2007, personal communication).

⁴ Information in brackets, [. . .], has been inserted by the author.

⁵ The Fields Medal, the so-called Nobel Prize for mathematicians, was awarded to Kodaira in 1954 for his major contribution to the theory of harmonic integrals. It was the first Fields Medal given to a mathematician in Japan.

inspired research on the jet stream and hemispheric circulation. And, of course, a momentous change in the scientific environment was produced by the computer—labeled the “great mutation” by Edward Lorenz (Lorenz 1996).

The work that most stimulated students at the Geophysical Institute was Jules Charney’s paper on baroclinic instability (Charney 1947)—the dynamic underpinning for midlatitude cyclone development. Although published in 1947, the students and the chair professor of meteorology, Shigekata Syono, only became aware of it two years later. In those early years of the occupation, a two-year delay in the transmission of information from the West was typical. Journals, if available at all, were housed at the Hibiya Library in the heart of Tokyo near General Douglas MacArthur’s headquarters. And even if a journal could be located, obtaining photocopies of an article was a nontrivial task.

3. Connection to America

Without exaggeration, it is fair to say that graduate students at the Geophysical Institute were separated from the worldwide centers of research activity in meteorology. A ray of hope for connectivity came with Kanzaburo Gambo’s invitation to join the Electronic Computer Project at Princeton University’s Institute for Advanced Study (IAS). Gambo received his B.S. degree in physics from the University of Tokyo in 1945, and was immediately hired as an assistant, Faculty of Science; in today’s terminology, he was a research associate. Inspired by Charney’s work mentioned above, Gambo began his own study of baroclinic disturbances and established a modified criterion for the development of cyclones (Gambo 1950). This note was sent to Charney, and Gambo received a swift and encouraging reply with an offer to join the celebrated group at the IAS. These researchers had recently made two 24-h numerical weather predictions of transient features of the large-scale flow over the United States (Charney et al. 1950; reviewed in Platzman 1979).

During his stay at Princeton from October 1952 to January 1954, Gambo frequently sent detailed reports to the staff and students at the Geophysical Institute. These reports stimulated the formation of a numerical weather prediction (NWP) group under the leadership of Syono (and Gambo after he returned to Japan in 1954). There were approximately 25 members of this group, which included JMA employees as well as the graduate students. They were, of course, without the benefit of computational power. In the absence of high-speed computation, the motivated students made use of



FIG. 1. Faculty, staff, and students of the Geophysical Institute, University of Tokyo, in the early 1950s. Numbering from left to right in each row, the following individuals have been identified. Front row: 1) Mariko Ookkubo, secretary; 2) [?] Koide, student; 3) Yoshi Sasaki, student; and 4) Dr. Yoshinari Nakagawa, research assistant. Back row: 1) unknown; 2) Dr. Kanzaburo Gambo, research associate; 3) Chair Professor Shigekata Syono; and 4) Dr. [?] Tao, research associate. (Courtesy of A. Kasahara.)

desk calculators, but more often, they performed the integration of the governing equations via graphical methods (see Fjørtoft 1952).

The students who were at the Geophysical Institute in the early 1950s view Gambo as a scientific hero. His altruistic efforts to build a bridge to the West are always at the forefront of their thoughts as they retrospectively examine their own careers. A photo of Gambo, Syono, and Sasaki is found in Fig. 1.

4. Track prediction of typhoons

Among the first contributions that came from the NWP group in Tokyo was a paper by Sasaki and Miyakoda on the track forecasting of typhoons (Sasaki and Miyakoda 1954). It is the first known publication on numerical prediction of typhoon/hurricane tracks.⁶ Twenty-four-hour track predictions were made by partitioning the geopotential field at 700 mb into a steering current (synoptic-scale flow component) and a smaller-scale circular vortex to fit the geopotential in the immediate vicinity of the typhoon. The structure of the circular vortex followed Ted Fujita’s empirical formula (Fujita 1952).⁷ By extracting the large-scale component from the analyzed geopotential and combining this with

⁶ See Demaria (1996) for a history of hurricane track forecasting in the Atlantic basin.

⁷ At that time, Fujita was a professor of physics at the Kyushu Institute of Technology.

the 24-h prediction of the geopotential tendency (found from the graphical solution to the barotropic vorticity equation), the typhoon's motion could be determined. Several cases were tested with excellent results. This experience also brought Sasaki into contact with operational forecasting at JMA. It would prove to be an invaluable experience in his later quest to devise an objective analysis scheme for numerical prediction.

Sasaki and Miyakoda presented their paper at the United Nations Educational, Scientific and Cultural Organization (UNESCO) Symposium on Typhoons, held in Tokyo on 9–12 November 1954. Miyakoda has a vivid memory of events surrounding this presentation. As he recalled:

We [Sasaki and Miyakoda] worked together in the university on developing a method of typhoon tracks and presented our work at the international meeting in Tokyo. In that meeting, the scientists attending from US were H. [Herbert] Riehl, [Father Charles] Deppermann [S. J.],⁸ Bob Simpson, [Leon] Sherman, possibly J. [Jerome] Namias. This is the first international meeting after the Pacific War. This method had been used in Japan Meteorological Agency for more than 10 years since then.

(K. Miyakoda 2007, personal communication)

Beyond its influence on operations at JMA, the work of Sasaki and Miyakoda stimulated research in the United States—first at the University of Chicago and later at the U.S. Weather Bureau (USWB). Akira Kasahara, a postdoctoral fellow at the University of Chicago in the late 1950s, recalls this work:

When I joined Platzman's research project at University of Chicago in 1956, I decided to extend the steering flow method of Sasaki and Miyakoda (1954) to suit for the numerical prediction of hurricanes using the electronic computer at the Joint Numerical Weather Prediction Unit in Suitland, Maryland [see Kasahara (1957)]. Working with George Platzman, Gene Birchfield, and Robert Jones, we developed a series of numerical prediction models of hurricane tracks. Lester Hubert of the U.S. Weather Bureau took one of our prediction schemes, tested [it] operationally and published an article (Hubert 1959).

(A. Kasahara 2007, personal communication)

5. Sasaki's dissertation

As an undergraduate in physics, Sasaki was captivated by the subject of variational mechanics, that post-Newtonian branch of analytical mechanics developed

by Leonhard Euler, Joseph Louis Lagrange, and William Rowan Hamilton. Historically, this subject has been found so esthetically pleasing and utilitarian that words like “poetic,” “Shakespearian,” and “conspicuously practical” are used to describe it (Bell 1937; Lanczos 1970).⁹ Sasaki had seen its application in quantum mechanics and the theory of relativity, but as he said, “I was fascinated to find whether or not there exists variational principle for meteorological phenomenon” (Y. Sasaki 2007, personal communication). This research topic was far afield from other efforts at the Geophysical Institute, but “Professor Syono encouraged my ‘lone-wolf’ approach and asked me to make informal presentations to him” (Y. Sasaki 2007, personal communication). As elaborated upon by Sasaki:

There was minimal literature on the use of variational methods in fluids—some in Horace Lamb's book on hydrodynamics [Lamb 1932] and some applications to irrotational fluids in Bateman's book [Harry Bateman, Caltech mathematician and aerodynamicist; Bateman (1932)]. I had been working with Miyakoda on typhoon tracking and wanted to apply these variational ideas to a typhoon that would be idealized as a vortex. The Clebsch transformation [Clebsch 1857] is the key to this problem, and Syono wanted me to discuss the physical basis for the transformation. I learned a lot from these discussions, and I made good progress. The results from the work formed the major part of my dissertation research, which was published in 1955 [Sasaki 1955].

(Y. Sasaki 1990, personal communication)

We should not be surprised that Sasaki found stimulation from Bateman, for he was a true disciple of Hamilton, to the variational principle, and the author of the monumental treatise *Partial Differential Equations of Mathematical Physics* (Bateman 1932).

Sasaki's development of these ideas for geophysical fluid dynamics was a valuable extension of the work accomplished by Bateman (1932, section 2.52); that is, Bateman's case of irrotational flow in fixed coordinates was expanded to include the generalized hydrothermodynamic equations on a rotating earth with application to the problem of tracking the vorticity associated with a typhoon. Again, in the absence of digital computers, Sasaki integrated the governing equations via graphical methods. Several years later, without knowledge of Sasaki's work, the noted Scripps physicist–oceanographer Carl Eckart developed similar equations for the ocean–atmosphere system (Eckart 1960a,b).

⁹ In Eric Temple Bell's history, these superlative words are found in his discussion of Lagrange's work. This lofty language is found in chapter 10 of Cornelius Lanczos's book.

⁸ Society of Jesus (Jesuits), an order of the Catholic Church.

During his final year of graduate study, Sasaki began to consider the full scope of deterministic NWP as it existed in the mid-1950s—initialization and prediction with quasigeostrophic models. His experience with track prediction of typhoons over the western Pacific Ocean made him aware of the challenges that come with analysis in data-sparse regions. How to couple the observations via dynamical constraints became a central concern and led him to consider weather map analysis based on the calculus of variations.

6. Variational view of objective analysis amid pragmatism

In the presence of the necessarily pragmatic and computationally efficient objective analysis schemes that surfaced in response to the needs of NWP in the 1950s, there came two theoretical developments that were too computationally demanding for operational implementation in the 1950s. Nevertheless, these developments pointed to a promising future for operational analysis and prediction. The authors of these theoretical developments were Arnt Eliassen (Eliassen 1954) and Sasaki (Sasaki 1958). Both approaches rested on minimization of a least squares cost functional; accordingly, they are labeled variational methods. The fundamental difference in these variational methods is philosophical: one based on a deterministic view of the physical system (Sasaki 1958) and the other on a stochastic view (Eliassen 1954). Determinism implies that the future state of the system is controlled by the present state and the casual laws governing the system [historically linked to Laplace's "mechanistic" view of the universe (Bohm 1957)]. The stochastic process, on the other hand, implies an uncertain view of the universe where probability replaces certainty in the casual laws and observations. Eliassen built his analysis on this framework.¹⁰

To clarify the interplay between the optimal methods and operational-pragmatic schemes, we subdivide this section as follows: (a) operational schemes, (b) stochastic optimality (Eliassen-Gandin), (c) deterministic optimality (Sasaki), and (d) deterministic versus stochastic analysis.

a. Operational schemes (1940s–1960s)

1) POLYNOMIAL FIT (PANOFSKY AND GILCHREST-CRESSMAN)

Hans Panofsky, professor of meteorology at New York University in the late 1940s, devised the first ob-

jective analysis scheme intended for operational use (Panofsky 1949). From knowledge of research associated with the Electronic Computer Project, Panofsky realized that NWP was a likely replacement for subjective forecasting—to wit, an objective analysis method would be required. He took a meaningful first step—an analysis based on the least squares fit of a polynomial (third degree in his case) to the upper-air observations of geopotential and wind. The functional to be minimized consisted of 1) a least squares fit of the analysis to the observed geopotential and 2) a least squares fit of the gradient of the analysis to the “observed gradient of geopotential” (via the observed wind and geostrophy). Since a third-degree polynomial contains 10 coefficients, the fit is achieved through the use of at least 10 observations (the requisite number). Yet, there is an advantage to using more than the requisite number to incorporate smoothing into the analyzed field (essentially, reduction of the random error in observations). The development of governing equations for analysis is straightforward, but inversion of the associated matrix was far from straightforward and certainly inefficient by today's standards. (Panofsky's extended discussion concerning the matrix inversion problem gives the reader an appreciation of obstacles facing objective analysis in the early 1950s.) Aside from difficulties with the inversion process, the polynomial was unable to account for variations in the curvature of the typical midlatitude synoptic pattern over the wide expanse of the United States. Thus, the United States was divided into subsections where separate polynomial fits were constructed. But the associated problem of matching solutions at the boundaries of these subdomains sounded the death knell for this approach.

In an effort to overcome the weakness of Panofsky's method, theoretician Bruce Gilchrist teamed with George Cressman, Rossby protégé and head of the Joint Numerical Weather Prediction Unit (JNWPU). They developed a scheme that fit a second-degree polynomial to the observations around each grid point in the NWP domain (Gilchrist and Cressman 1954). The local fit around each point faithfully reconstructed the features of the large-scale geopotential pattern, but the machine computations for the inversion process (the order of several hundred every 12 h) exceeded the strict time requirements for operational objective analysis in the 1950s–1960s.

2) INCREMENTAL ANALYSIS WITH BACKGROUND (BERGTHÓRSSON-DÖÖS AND CRESSMAN)

The first computer-based operational analysis scheme was developed in Sweden under the watchful eye of Carl-Gustaf Rossby (A. Wiin-Nielsen 1993, per-

¹⁰ Daley (1991, p. 98) gives a lucid description of stochastic processes.

sonal communication). Rossby chose two protégés to lead this charge: the theoretically inclined Bo Döös and the Icelandic forecaster Páll Bergthórsson (see Fig. 2). Output of the analysis scheme would provide the initial height field for the quasigeostrophic model developed by another team of researchers at “Rossby’s Institute,” the International Institute of Meteorology at Stockholm University (Wiin-Nielsen 1991). Bergthórsson recalls the situation:

I was a weather forecaster in Reykjavik in Iceland when Rossby invited me, as well as scientists from several countries, to stay at his institute in the University of Stockholm. Rossby told me and Bo to figure out how to derive a height analysis untouched by human hand. My cooperation with Bo was very pleasant. He was a very creative scientist and an eminent programmer. I had, on the other hand, been a student of the great master of weather map analysis, the Swedish professor Tor Bergeron, and my experience in the forecasting center at Reykjavik was probably of some value. We decided to imitate the manual analysis in a way. The latest NWP map should be used to ensure continuity of the maps, always keeping climatology in mind. The forecast should then be corrected with the available observations of wind and height. To find the correction of a certain grid point, all surrounding observations were used, giving them weights according to their distance from the grid point. The weighting as a function of distance was determined statistically on analyzed maps. An iteration of the analysis proved to be of some value.

(P. Bergthórsson 2007, personal communication)

In Bergthórsson and Döös (1955, hereafter BD), the objective analysis of 500-mb height followed the principles outlined by Bergthórsson. The region of interest spanned the North Atlantic Ocean and the bounding continental areas (western Europe and eastern Canada). The spacing of points was uniform with a separation interval of ~ 300 km. In the BD scheme, a “background” geopotential height was found at every grid point within the region.¹¹ This background information was a linear combination of 12-h forecast and climatology where the weights on forecast and climatology were inversely proportional to the square root of their respective error variances.

The background field was interpolated to locations of all observations, and increments–D values (observation – background) were found at observation sites. To the eye of a forecaster like Bergthórsson, the pattern of D values was often indicative of a phase error in the forecasted synoptic-scale trough–ridge system. That is,

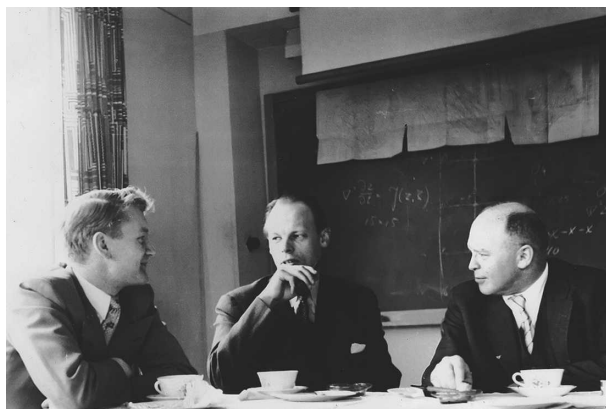


FIG. 2. Páll Bergthórsson, Bo Döös, and Carl Rossby (left to right) in conversation at Stockholm University (ca 1955). (Courtesy of P. Bergthórsson.)

if the forecasted trough–ridge system moved too slowly or too fast, there would be a juxtaposition of a “+D pattern” and a “–D pattern” on the scale of the system. Distance-dependent weighting of these increments around each grid point was central to the scheme. As in Panofsky’s method, wind observations were used to structure gradients of geopotential. [See section 2 of Bergthórsson and Döös (1955) for details.]

Shortly after development of the BD method, Cressman devised a closely related scheme (Cressman 1959). As stated in Cressman (1959), “The system described below is based essentially on the general method described by Bergthórsson and Döös [2] and resembles to some extent the method recently reported by Haug [6].”¹² Odd Haug’s appealing approach was used in operations at the Norwegian Weather Service (Norwegian Meteorological Institute). This approach incorporated a novel smoothing process that removed nonsystematic errors by repeatedly passing a linear finite-difference operator through the network of grid points. Haug credited Ragnar Fjørtoft, then director of the Norwegian Meteorological Institute, with the idea.

Although Cressman’s method was iterative, it was computationally efficient. He incorporated simple analytic functions for the weighting that mimicked the statistical bell-shaped weighting functions of BD. Further, the background field was the forecast. The scheme rested on the principle that larger- and smaller-scale spatial features in the field could be sequentially constructed. The first iteration was designed to capture the large-scale component while subsequent iterations built the smaller-scale structure into the analyzed field.

¹¹ “Background” is sometimes referred to as “first guess,” but the most appropriate nomenclature is *prior* from Bayesian theory.

¹² References [2] and [6] in the quotation appear as Bergthórsson and Döös (1955) and Haug (1959) in this paper.

Let us mathematically describe the procedure. [See Lewis et al. (2006) for details.] The observations on the 500-mb surface are represented by $\mathbf{z} \in \mathbf{R}^m$, a real column vector of size m , and the field to be analyzed is represented by $\mathbf{x} \in \mathbf{R}^n$. Since the procedure is iterative, subscript k indicates the iteration number; that is, \mathbf{x}_k represents the field at the k th iteration. The algorithm is given by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{QW}(\mathbf{z} - \mathbf{Hx}_k), \quad k = 0, 1, 2, \dots, \quad (1)$$

where $\mathbf{x}_0 = \mathbf{x}_B$ denotes the background field, $\mathbf{W} \in \mathbf{R}^{n \times m}$ is a real $n \times m$ matrix called the weight matrix, $\mathbf{Q} \in \mathbf{R}^{n \times m}$ is a diagonal matrix with normalizing constants across its diagonal, and $\mathbf{H} \in \mathbf{R}^{n \times m}$ is the interpolation matrix. The diagonal elements of \mathbf{Q}^{-1} are

$$Q_{ii}^{-1} = \sum_{j=1}^m W_{ij} \quad \text{and} \quad (2)$$

$$W_{ij} = \frac{(d_k^2 - r_{ij}^2)}{(d_k^2 + r_{ij}^2)} \quad \text{for } r_{ij} \leq d_k; \quad (3)$$

$$W_{ij} = 0, \quad \text{otherwise.}$$

The distance between the grid point (indexed by i and j) and the observation location is denoted by r_{ij} , and d_k is the radius of influence on the k th iteration. The magnitudes of the radii are sequentially reduced. On the first iteration, the largest radius is used and it is designed to capture the large spatial scale of the field. The result stemming from this first iteration serves as input to the second iteration. With a smaller radius of influence, this second iteration adds smaller-scale detail to the analysis, etc. Operational implementation of Cressman’s method at the National Meteorological Center [NMC, now known as the National Centers for Environmental Prediction (NCEP)] in the late 1950s used a forecast horizon of 12 h and the following radii of influence: 4.75, 3.60, 2.20, and 1.80, all multiples of the mesh length (≈ 281 km).

b. Stochastic optimality (Eliassen–Gandin)

A stochastic or statistical-based approach labeled optimum interpolation (OI) was introduced into meteorology by the mid-1950s. Arnt Eliassen (Fig. 3) pioneered this approach as found in a “gray” paper, the appendix to a technical report from Oslo University (Eliassen 1954).¹³

¹³ “Optimum interpolation” was introduced into the English-speaking meteorological literature in the early 1960s following the translation of Gandin (1965) from Russian to English (A. Lorenc 2007, personal communication). In Russian, the term implies that all statistical properties are known with absolute accuracy. “Statistical interpolation” (SI) became a more appropriate term due to the inevitable uncertainty of the statistical properties when applied to meteorological analysis (Rutherford 1972).



FIG. 3. Arnt Eliassen, in kimono, at the banquet for the First Conference on NWP in Tokyo, Japan (1960). (Courtesy of G. Platzman.)

Eliassen’s scientific heritage stems from Vilhelm Bjerknes, the venerated leader of the Bergen School who became professor of mechanics and theoretical physics at the University of Oslo in 1926. Eliassen’s mentor was Einar Höiland, a Carnegie assistant under Bjerknes. The Carnegie Institution’s grant that supported Höiland was awarded to Bjerknes following his delivery of a series of lectures at the institution in 1905. This yearly support lasted for nearly 35 years. Eliassen’s recollection of Höiland follows:

Vilhelm Bjerknes’s last Carnegie assistant (from 1935) and collaborator for many years was Einar Höiland. Together they held weekly seminars on meteorology, hydrodynamics, thermodynamics, and statistical physics which attracted inquisitive students. Both Ragnar Fjørtoft and the author of this article [Eliassen] were captured for meteorology in this way. After the war, Höiland became professor of hydro- and aerodynamics at Oslo. Like his teacher, Vilhelm Bjerknes, he too had a great talent for attracting and inspiring students . . . and Einar Höiland was the one who took the most interest in me; he would act as my advisor more than anyone else.

(Eliassen 1982)

Eliassen was granted his bachelor’s degree in 1939 and his doctorate in 1941 at the University of Oslo. He then

took a position as a meteorologist for the Norwegian Weather Service and remained there until the early 1950s. He was a major contributor to the Electronic Computer Project at Princeton in the late 1940s–early 1950s.

While a member of the World Meteorological Organization's (WMO) working committee on network design in 1956, he presented the stochastic objective scheme to committee members including Russian meteorologist Lev Gandin. As Eliassen stated in his oral history interview:

I had been there [WMO] working on a method to analyze the pressure field, or any other field, by utilizing the covariances between the various stations, and this was taken up by some people in the working group on the network, and how the observation of networks should be set up, so we had several meetings on this question. The method was developed further by several other people, in particular by Lev Gandin. (Eliassen 1990, p. 23)

Gandin's expansive development is found in his textbook that was translated by the Israel Program for Scientific Translations (IPST) in 1963. It became available worldwide in 1965 (Gandin 1965). An independently developed and little-known work by Canadian meteorologist Amos Eddy has much in common with the stochastic approach of Eliassen–Gandin (Eddy 1967). Eddy's work was inspired by his contact with noted MIT statistician George Wadsworth (A. Eddy 1969, personal communication).

In what follows, we describe the essential features of the OI scheme as developed by Eliassen (1954). The reader is referred to Lewis et al. (2006, chapter 34) for details on the derivation. For his study, Eliassen used upper-air data from 15 stations in northwestern Europe. These data were collected over the months of December and January for 6 yr (December 1948–January 1954).

Eliassen outlined the method by choosing an arbitrary grid point on the 500-mb surface in his domain of analysis. This grid point is identified by the subscript "0." The analysis problem reduces to finding the 500-mb height at 0 (at a particular time) by using observations from the surrounding 15 observation sites at this same time.

His first step is to find the mean value of the 500-mb height at each station over the months and years of interest (mentioned above). Thus, 96 observations go into the calculation of the mean 500-mb height at each station. These mean values define a 15-component vector, $\bar{\mathbf{z}}$. Then, at each station, he constructs a D value or *incremental observation vector*, $\tilde{\mathbf{z}} = (\mathbf{z} - \bar{\mathbf{z}})$. The observation vector is given by $\mathbf{z} = (\mathbf{z}' + \mathbf{v})$, where \mathbf{z}' is the

truth and \mathbf{v} the observational noise–error. The *covariance* of this observational noise is denoted by $\mathbf{R} \in \mathbf{R}^{n \times m}$. Eliassen also assumes that the mean value of the analysis at 0 is known and denoted by $\bar{\mathbf{x}}_0$. The unknown analysis at 0 is denoted by \mathbf{x}_0 , and the associated *analysis increment* is given by $\tilde{\mathbf{x}}_0 = (\mathbf{x}_0 - \bar{\mathbf{x}}_0)$.

Eliassen's basic tenet stipulates that the analysis increment at 0 can be expressed as a linear combination of the observation increments. That is,

$$\tilde{\mathbf{x}}_0 = \sum_{j=1}^m w_j \tilde{\mathbf{z}}_j = \mathbf{w}^T \tilde{\mathbf{z}}, \quad (4)$$

where m is the number of observations ($=15$), $\mathbf{w} \in \mathbf{R}^m$ is the unknown weight vector, and T denotes the transpose. The unknown weights are found by the minimum variance technique, that is, minimizing the mean-square error of the analysis increment, $f(\mathbf{w})$, given by

$$f(\mathbf{w}) = E[\tilde{\mathbf{x}}_0 - \mathbf{w}^T \tilde{\mathbf{z}}]^2, \quad (5)$$

where $E[\dots]$ denotes an ensemble average. The optimal \mathbf{w} is a function of $\mathbf{B} = E[\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T] \in \mathbf{R}^{n \times m}$, $\mathbf{d}_{j0} = E[\tilde{\mathbf{z}}_j \tilde{\mathbf{x}}_0] \in \mathbf{R}^m$, and the observation error covariance, \mathbf{R} .

If we let

$$\mathbf{d}_{*0} = (d_{10}, d_{20}, \dots, d_{m0})^T, \quad (6)$$

then the optimal weights are found from the solution of

$$\mathbf{w} = (\mathbf{B} + \mathbf{R})^{-1} \mathbf{d}_{*0}. \quad (7)$$

Eliassen (1954) makes no mention of earlier contributions that may have influenced him. He was familiar with statistical physics and that may have been sufficient background for this work. His approach has similarity to the classic treatises of Kolmogorov (1941) and Wiener (1947) in random function theory [reviewed in Yaglom (1962)].

Although Eliassen's analysis was univariate, Gandin extended it to multivariate analysis under the geostrophic constraint (Gandin 1965). With the forecast serving as the background instead of the long-term temporal mean values, Environment Canada and NMC used Gandin's geostrophic constraint form of OI as their midlatitude operational analysis in the mid-1970s (Rutherford 1972; Bergman 1975; review by Schlatter et al. 1976). By the early 1980s, Andrew Lorenc was able to extend the OI concept to a global domain at the European Centre for Medium-Range Weather Forecasts (ECMWF) (Lorenc 1981)—the pinnacle of OI in operations. Shortly thereafter, Arne Bratseth developed the very appealing version of OI based on successive corrections [Bratseth (1986)].

When the forecast is used as the background, Eliassen’s scheme takes the form

$$\mathbf{x} - \mathbf{x}_B = \mathbf{K}(\mathbf{z} - \mathbf{H}\mathbf{x}_B), \tag{8}$$

where $\mathbf{K} \in \mathbf{R}^{n \times m}$ is the counterpart to \mathbf{w} , $(\mathbf{x} - \mathbf{x}_B)$ is the counterpart to $\tilde{\mathbf{x}}$, and $(\mathbf{z} - \mathbf{H}\mathbf{x}_B)$ is the counterpart to $\tilde{\mathbf{z}}$ (\mathbf{H} is the interpolation matrix). In this case, the optimal weight is given by

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}, \tag{9}$$

where \mathbf{B} represents the covariance matrix of the background (forecast) error. Although $(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})$ is rarely singular in practice (operations), “near singularity” can occur. Quoting Jim Purser, research meteorologist at the NCEP/Environmental Modeling Center (EMC):

Practical difficulties in numerical implementation of OI can occur when \mathbf{B} is an “almost singular” covariance, such as a Gaussian, and the components of \mathbf{R} are very tiny (corresponding to very precise measurements), especially when many of the measurements are clustered close together. At such small scales, the Gaussian shape tends to be too smooth—a better shape results from adding some amount of a small-scale covariance (which can still be Gaussian form when that shape is convenient). Also, all measurement should have finite \mathbf{R} since, even if they are certified as “precise,” the addition of the necessary “representative error”¹⁴ into the effective \mathbf{R} keeps this diagonal (or block diagonal) matrix sufficiently far from singularity.

(J. Purser 2007, personal communication)

c. *Deterministic optimality (Sasaki)*

1) **DIAGNOSTIC CONSTRAINTS**

Sasaki appreciated the need for efficient computer-generated analyses in the operational environment, yet from a more dynamical perspective, he realized that there was no guarantee that the operationally produced analyses would satisfy the zeroth-order constraint for the quasigeostrophic model—that is, the geostrophic wind law. Generally there would be a discrepancy between the analyzed horizontal gradient of the geopotential and the horizontal gradient based on geostrophy and the observed wind. How to use both the wind and geopotential observations in accordance with their accuracy in order to guarantee a dynamically consistent set of analyses was at the heart of his thought and contribution.

In concert with his training in theoretical physics, Sasaki posed the problem in a continuous-space frame as opposed to the discrete frame used by Eliassen. The variational analysis was derived for two separate diagnostic constraints: the geostrophic wind law and the balance equation—both stemming from separate scaling of the divergence equation (see Thompson 1980).

Under the constraint of geostrophy, namely,

$$\mathbf{v} = \frac{1}{f} \frac{\partial \phi}{\partial x}, \quad u = -\frac{1}{f} \frac{\partial \phi}{\partial y}, \tag{10}$$

where (u, v) are the wind components in the $[x(\text{positive east}), y(\text{positive north})]$ coordinate frame, f is the Coriolis parameter, and ϕ is the geopotential. The functional to be minimized is given by

$$J = \iint_{x,y} \left\{ \frac{1}{\sigma_\phi^2} (\phi - \hat{\phi})^2 + \frac{1}{\sigma_v^2} [(u - \hat{u})^2 + (v - \hat{v})^2] \right\} dx dy, \tag{11}$$

where $\hat{\phi}$, \hat{u} , and \hat{v} are the continuously distributed observations, and σ_v^2 , σ_ϕ^2 are the error variances of the wind and geopotential, respectively. The associated Euler–Lagrange (EL) equation is

$$\frac{1}{f} \left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} \right) - wf\phi = -wf\hat{\phi} + \hat{\varsigma}, \tag{12}$$

where $\hat{\varsigma}$ is the observed vorticity, $[(\partial\hat{v}/\partial x) - (\partial\hat{u}/\partial y)]$, and w is the weight ratio $[w = (\sigma_v^2/\sigma_\phi^2)]$.

Natural boundary conditions for this problem allow two choices: Dirichlet conditions (geopotential on boundaries given by observations of geopotential) or Neumann conditions (normal derivative of geopotential given on the boundary via geostrophy and the observed winds).

There is an esthetic beauty to these analysis equations. The observed winds and geopotential contribute to the solution in accord with their relative accuracy (via the ratio of weights). When only wind observations are available ($w = 0$), the governing EL equation reduces to

$$\frac{1}{f} \left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} \right) = \hat{\varsigma}, \tag{13}$$

under Neumann boundary conditions—governing equation and boundary conditions totally consistent with geostrophy.

Under the balance equation constraint, the EL equation is more complicated (and nonlinear) as expected,

¹⁴ These are errors associated with meteorological scales that cannot be resolved by the observation network.

yet Sasaki exhibits an adroitness with mathematical physics and solves the governing analysis equation in terms of Green's functions. The associated integral equation is solved by successive corrections–approximations. This exposition is a mathematical tour de force.

2) DYNAMIC CONSTRAINTS

With the hope and promise of longer-range weather prediction (beyond several days) that stemmed from Phillips's successful numerical experiment linked to hemispheric circulation (Phillips 1956), advances in general circulation modeling using more-realistic models, improved high-speed computation [e.g., Smagorinsky (1963)], and with the politico-scientific thrust that was aimed at the acquisition of global observations [addressed in Smagorinsky (1978)], Sasaki rekindled his interest in objective analysis and expanded his view to encompass the temporal distribution of observations. His expanded view of data assimilation is found in a trilogy of papers published in the December 1970 issue of the *Monthly Weather Review* (Sasaki 1970a–c). This development was made within the discrete frame of reference, and he labeled the process numerical variational analysis. A photo of Sasaki shortly after completing this work is shown in Fig. 4.

A succinct mathematical description of the methodology follows. Let $\mathbf{z}_k \in \mathbf{R}^m$ for $k = 1, 2, \dots, N$ be the set of N observations obtained at N distinct times. Assume the model (generally nonlinear) takes the form

$$\mathbf{x}_{k+1} = \mathbf{M}(\mathbf{x}_k), \quad (14)$$

where $\mathbf{M}:\mathbf{R}^n \rightarrow \mathbf{R}^n$ and $\mathbf{M}(x) = [M_1(x), M_2(x), \dots, M_n(x)]^T$ and $\mathbf{M}[\cdot]$ denotes the discrete form of the model operator that maps the state at time k onto the state at time $(k + 1)$. When the model is linear, Eq. (14) takes the form

$$\mathbf{x}_{k+1} = \mathbf{M}\mathbf{x}_k, \quad (15)$$

where $\mathbf{M} \in \mathbf{R}^{n \times n}$ is an $n \times n$ real matrix. For the deterministic model, specification of the initial state $\mathbf{x}_0 \in \mathbf{R}^n$ determines the sequence of future states ($\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$). The problem reduces to finding \mathbf{x}_0 that minimizes

$$J(\mathbf{x}_0) = \frac{1}{2} \sum_{k=1}^N [\mathbf{z}_k - \mathbf{h}(\mathbf{x}_k)]^T \mathbf{R}_k^{-1} [\mathbf{z}_k - \mathbf{h}(\mathbf{x}_k)], \quad (16)$$

where $\mathbf{R}_k^{-1} \in \mathbf{R}^{m \times m}$ is the inverse of the covariance of the observational errors–noise at time k (temporally uncorrelated), and $\mathbf{h}:\mathbf{R}^n \rightarrow \mathbf{R}^m$ with $\mathbf{h}(\mathbf{x}) = [h_1(x), h_2(x), \dots, h_m(x)]^T$, the mapping that connects the model state at time k with the observation at the same



FIG. 4. Sasaki at his desk in the Engineering Laboratory Building at the University of Oklahoma (May 1970). (Courtesy of J. Heimbach.)

time and is labeled the “model counterpart to observation.”

The problem is solved by converting the constrained problem to an unconstrained problem via the use of the Lagrange's undetermined multipliers. Let $\mathbf{L}:(\mathbf{R}^n)^N \times \mathbf{R}^n \rightarrow \mathbf{R}$, where

$$\mathbf{L}(\boldsymbol{\lambda}, \mathbf{x}_0) = \mathbf{J}(\mathbf{x}_0) + \sum_{k=1}^N \boldsymbol{\lambda}_k^T [\mathbf{x}_{k+1} - \mathbf{M}(\mathbf{x}_k)], \quad (17)$$

with $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^T, \boldsymbol{\lambda}_2^T, \dots, \boldsymbol{\lambda}_m^T)^T$ and $\boldsymbol{\lambda}_k \in \mathbf{R}^n$ for $1 \leq k \leq N$. In this circumstance, we seek the multipliers $\boldsymbol{\lambda}_k \in \mathbf{R}^n$, $1 \leq k \leq N$, and the initial conditions $\mathbf{x}_0 \in \mathbf{R}^n$ that minimize $\mathbf{L}(\boldsymbol{\lambda}, \mathbf{x}_0)$. The optimal state is found by solving the discrete form of the EL equations.¹⁵

Sasaki's approach seems so straightforward in retrospect, yet the magnitude of the contribution is ever so impressive in light of the fact that such a methodology had never before been used in continuum mechanics. Most fundamentally, Sasaki's method is tied to Gauss's monumental work in 1801 when the method of least squares under a constraint was developed (Gauss 1809). In Gauss's classic work, Kepler's laws of celestial mechanics were used as constraints. Sasaki was un-

¹⁵ By the early 1980s, functionals such as Eq. (17) were minimized by using adjoint methods in conjunction with minimization algorithms (discussed in section 7).

aware of Gauss's work on this subject (Y. Sasaki 2007, personal communication).

In 1971, an upper-air analysis based on Sasaki's method was implemented at the U.S. Navy's Fleet Numerical Weather Center (FNWC; Lewis 1972). The generalized nonlinear thermal wind equation, applicable over the globe, was used as the constraint. This optimal upper-air analysis remained operational at FNWC [now known as the Fleet Numerical Meteorological and Oceanographic Center (FNMOC)] until the mid-1990s (E. Barker 2004, personal communication).

d. Deterministic versus stochastic analysis

Although Sasaki developed a "weak constraint" or penalty function form of the variational analysis, that is, a form that allowed some departure from an exact or strong constraint (Sasaki 1970a), his approach fundamentally follows the Gaussian tradition—determination of initial conditions that minimize the least squares fit of a model to observations. The weights are specified in terms of their general knowledge of the observational errors. The algorithmic structure is "offline"; that is, a historical set of observations is treated in a collective fashion. This form of analysis is especially conducive to reanalysis over a time window that couples all observations with the constraints.

The "online" or sequential structure of the stochastic objective analysis makes it appealing for operations. The background implicitly carries information from the past to the present. Further, explicit account of the spatial aspects of the background errors' covariances give it advantages over the Sasaki scheme. As viewed by Purser:

Within the Gandin paradigm, there exists one particular model for the background covariance which is formally equivalent to the filtered characterization of errors implied by Sasaki's variational principle, but this would be a very special and contrived covariance model (and would often be somewhat singular at zero spatial separation). When I first encountered Sasaki methods at the Met Office [U.K. Meteorological Office], they were only used to blend separate univariate gridded analysis (with geostrophy as a weak constraint) because I think they led to singularities forming at observation positions if we ever tried to use the spot measurements in these schemes directly. This is the kind of problem that revealed a weakness or limitation in the Sasaki method as it was then presented. Incorporation of higher derivatives in the spatial part of the variational scheme, as was later proposed by Wahba and Wendelberger (1980), smooths away such singularities and, formally at least, allows the Sasaki methods to apply to the measurements directly without resulting in spiky analyses. In this case, the

Green's functions of the higher-order elliptic operators in the Wahba and Wendelberger variants of Sasaki's method can be interpreted as the covariance operators of the corresponding equivalent Gandin OI schemes . . . a disadvantage of the OI scheme is determination of background error covariance for wind/mass relationships such as the balance equation. This constraint is not local as in the case of geostrophy, and it is not clear how one would set about formulating the multivariate covariance in such a case.

(J. Purser 2007, personal communication)

7. Present-day data assimilation

As stated earlier, there arose a wave of politico-scientific support for a global research program in the 1960s. Such a program was expected to build on the success of the International Geophysical Year (1957–1958) and promote peaceful use of space and space technology. Central to this space technology was the planned use of satellites for meteorological purposes (Smagorinsky 1971, 1978). Through the adoption of United Nations Resolutions 1721 and 1802 (in the year 1962), the international World Weather Watch program was created. Its principal component, the Global Atmospheric Research Program (GARP), began in 1968 and among its major thrusts was the routine collection of observations from satellites that circled the globe.

How to use the satellite measurements of upwelling radiation from the earth and its atmosphere—certainly not a model variable—became the question that faced data assimilators. Lewis Kaplan laid out a theoretical framework for temperature retrieval from radiance measurements in the immediate post-Sputnik period (Kaplan 1959). This plan rested on the use of infrared radiation measurements in the 15- μm band of CO_2 . Planck's radiation law used in conjunction with these measurements is the basis for reconstruction. However, implementation of the plan is complicated. The complication arises from the ill-posed nature of the problem. Planck's law takes the form of a Fredholm integral equation in this case. The inversion suffers from the strong overlap of weighting functions inside the integral. In short, there is no unique temperature profile that matches the observed radiances. To overcome the ill-posed nature of this problem, information from a *prior* (the forecast) is needed. Thus, the new era of data assimilation would come to depend on a Bayesian approach—an approach where statistical properties of the prior in conjunction with the statistical structure of the observational errors is used to generate a *posterior* estimate of the system state. Fundamentals of the Bayesian approach applied to meteorological data assimilation

are found in Lorenc (1986) and Lewis et al. (2006, chapter 20).¹⁶

The modern methods of meteorological data assimilation fall into two general categories labeled 3DVAR (for three-dimensional variational data assimilation, or 3D spatial analysis) and 4DVAR (for spatial and temporal analysis). We outline the principal foundations of these methods and the associated algorithms, but we refer the reader to Lorenc (1986), Daley (1991), and Rabier and Liu (2003) for important summaries.

a. 3DVAR

We consider two components of the functional to be minimized: 1) the square of the departure from the observations, $\mathbf{J}_0(\mathbf{x})$, and 2) the square of the departure from the background, $\mathbf{J}_B(\mathbf{x})$. The components of the functional are given by

$$\mathbf{J}_0(\mathbf{x}) = \frac{1}{2} [\mathbf{z} - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{z} - \mathbf{h}(\mathbf{x})] \quad (18)$$

and

$$\mathbf{J}_B(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}_B)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_B), \quad (19)$$

where our notation is consistent with that found in section 6 of this paper: \mathbf{B} is the covariance matrix of the background error and $\mathbf{h}(\mathbf{x})$ is the nonlinear observation operator. The combined functional is given by

$$\mathbf{J}(\mathbf{x}) = \mathbf{J}_0(\mathbf{x}) + \mathbf{J}_B(\mathbf{x}). \quad (20)$$

From Bayesian principles, maximizing the posterior probability of the analyzed system state is equivalent to the minimization of Eq. (20). To simplify the problem, $\mathbf{h}(\mathbf{x})$ is often approximated by first-order Taylor expansions around \mathbf{x}_B . That is, the problem is linearized and the solution is found iteratively. [See Lorenc (1986) and Lewis et al. (2006).] The first operational 3DVAR scheme was developed at NMC in the early 1990s (Parish and Derber 1992).

When the linearized form of $\mathbf{h}(\mathbf{x})$ is used, the solution reduces to that found from OI; namely,

$$\mathbf{x} = \mathbf{x}_B + \mathbf{B}\mathbf{H}(\mathbf{H}\mathbf{B}^T + \mathbf{R})^{-1}(\mathbf{z} - \mathbf{H}\mathbf{x}_B). \quad (21)$$

In this form, 3DVAR bears a strong resemblance to the Kalman filter method of data assimilation, a standard in engineering since the mid-1960s (Kalman 1960; Swirling 1959). A discussion of the similarities and differences between OI-3DVAR and Kalman filtering is

found in Ghil et al. (1981) and Lewis et al. (2006). The recently published textbook by Evensen (2007) and the review article by Lakshminarayanan and Stensrud (2008, unpublished manuscript) also explore the applications of ensemble Kalman filtering to meteorology.

b. 4DVAR

This formulation can be viewed as a straightforward extension of Sasaki's deterministic method as found in section 6c(2). The extension comes in the form of a term added to the cost functional—a term that fits the background to the analysis. The complete functional assumes the form

$$\mathbf{J}(\mathbf{x}_0) = \frac{1}{2} (\mathbf{x}_0 - \mathbf{x}_B)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}) + \frac{1}{2} \sum_{k=1}^N [\mathbf{h}_k(\mathbf{x}_k) - \mathbf{z}_k]^T \mathbf{R}_k^{-1} [\mathbf{h}_k(\mathbf{x}_k) - \mathbf{z}_k]. \quad (22)$$

Since we assume a deterministic model, all \mathbf{x}_k are functions of \mathbf{x}_0 and the model operator $\mathbf{M}(\mathbf{x})$ [see Eq. (14)]. Substituting the dynamical constraint into Eq. (22) converts the constrained minimization problem to an unconstrained problem; the initial condition \mathbf{x}_0 is the unknown. The solution can be efficiently found by using an adjoint method. The use of an adjoint model and a minimization algorithm obviates the need to solve the associated Euler-Lagrange equations—a much more demanding and often impossible task. The adjoint model theory for meteorological data assimilation was developed by LeDimet and Talagrand (1986) and used in practice by Lewis and Derber (1985) and Talagrand and Courtier (1987). Independently, Thacker and Long (1988) developed this approach and applied it to physical oceanography.

The first operational 4DVAR scheme was implemented at ECMWF in 1996 (Rabier et al. 2000). Subsequently, four other major operational weather prediction centers implemented 4DVAR: in order of implementation, France, the United Kingdom, Canada, and Japan. As stated in Rabier (2005), "Although much experience has been gained lately in 4DVAR, the understanding of the benefits brought about by this scheme is still an active field of research." A stimulating discussion on the relative merits of 3DVAR and 4DVAR may be found in Lorenc and Rawlins (2005). There is evidence that 4DVAR does not outperform 3DVAR on all counts.

8. Epilogue

Data assimilation, especially as viewed within the context of initializing dynamical weather prediction

¹⁶ Jim Purser is credited with advocating for the Bayesian approach to meteorological data assimilation at the British Meteorological Office in the late-1970s (see Purser 1984; A. Lorenc 2007, personal communication).

models, continues to present challenging problems. Compared to the dynamical models of the 1950s–1960s, modern-day models exhibit a level or degree of complexity and intricacy that rivals L. F. Richardson's phantasmagoric view of the atmosphere and its governing laws (Richardson 1922; Lynch 2006).¹⁷ And the observations are immensely more plentiful than in the early days of NWP. But with the increased number of observations comes difficulty—a difficulty linked to the “surrogate” nature of some observations. That is, observations like atmosphere–earth radiance and radar reflectivity from weather surveillance radars are complicated nonlinear functions of the model variables. And beyond the complexity of the models and the challenges of determining model counterparts to the observations, the model output must be verified against observations to determine the key statistical properties like the background error covariance. Yet, how can we creditably generate these statistics when the true state of the atmosphere is never known exactly? We use quantities whose error statistics are similar to those of the background error—for example, the differences between forecasts that verify at the same date–time, but that start at different initial times (See Fisher 2003). The task is daunting.

As we retrospectively examine the history of data assimilation in meteorology, it becomes clear that pragmatism has been inexorably linked with theoretical development. Rossby realized that success in operational objective analysis, the computer-generated analysis untouched by human hand, required the talents of a forecaster (Páll Bergthórsson) and a theoretically inclined researcher (Bo Döös). And during that decade of the 1950s when computer limitations dictated the pragmatic approach, Arnt Eliassen and Yoshi Sasaki ventured beyond the pragmatic and into the theoretical realm where they formulated two complementary forms of optimal analysis—one deterministic (Sasaki's) and the other stochastic (Eliassen's). Eliassen's motivation stemmed from an interest in network design while Sasaki sought to find a methodology that would combine observations under the constraint of dynamical law. These philosophically different approaches have afforded an expansive view of data assimilation in meteorology and they provide the underpinning for our current operational practice. We, the practitioners of data assimilation, owe much to these two meteorologists and to the other pioneers for their innovations in

this most challenging component of numerical weather prediction.

Acknowledgments. Oral histories and letters of reminiscence were received from the following individuals where “(…)”^O indicates oral history and “(…)”^L a letter of reminiscence with date(s) found in parentheses: Akio Arakawa^L (June 1992), Andrew Lorenc^L (December 2007), Edward Barker (conversation, June 2004), Kikuro Miyakoda^L (June 1992, May 2007), Páll Bergthórsson^L (June 2004, November 2007), Yoshimitsu Ogura^L (June 1992), Bo Döös^L (June 2004), Katsuyuki Ooyama^{O,L} (March 1992, September 1992), Amos Eddy (conversation, October 1969), Jim Purser^L (August 2007), Kanzaburo Gambo^L (June and September 1992), George Platzman^O (May 1990), Akira Kasahara^{O,L} (June 1991, May 2007), Yoshi Sasaki^{O,L} (May 1990, June 2007), Francois LeDimet (August 2007), and Olivier Talagrand (September 2007).

We extend our heartfelt thanks to each of these scientists.

We are especially grateful to Andrew Lorenc. He graciously gave us copies of lecture notes he used at the Atmospheric Data Assimilation Workshop, L'Aquila University, L'Aquila, Italy, in 2004. These notes contained important details on the historical development of data assimilation in meteorology.

David Schultz, editor for *MWR*, chose knowledgeable reviewers whose comments and suggestions went far to improve the manuscript. Jim Purser read the entire draft manuscript paying special attention to the mathematical details. His comments and suggestions were extremely valuable. Akira Kasahara and Togo Tsukahara carefully checked the manuscript for the accuracy of the recorded events in Japan (spelling of names, dates, times, and places). And for help in locating important historical papers, we commend Hiroshi Niino and Mashahito Ishihara (in Japan), and Frode Stordal, Trond Iversen, and Greta Krogvold (in Norway). Vicki Hall and Domagoj Podnar willingly aided in the electronic submission process.

Finally, we remember Tony Hollingsworth fondly and thank him for the stimulating discussions over the years that are central to this historical review. With his passing in 2007, we have lost a prized contributor and sterling representative in this field of meteorological data assimilation.

REFERENCES

- Bateman, H., 1932: *Partial Differential Equations of Mathematical Physics*. Cambridge University Press, 522 pp. [Reprinted by Dover Publication, 1944.]
- Bell, E., 1937: *Men of Mathematics*. Simon and Schuster, 592 pp.

¹⁷ In the introduction of Dover Publication's reprint of Richardson (1922), Sidney Chapman said, “He [Richardson] told me that he had put all that he knew into this book.”

- Bergman, K. H., 1979: Multivariate analysis of temperatures and winds using optimum interpolation. *Mon. Wea. Rev.*, **107**, 1423–1444.
- Bergthórsson, P., and B. Döös, 1955: Numerical weather map analysis. *Tellus*, **7**, 329–340.
- Bohm, D., 1957: *Causality and Chance in Modern Physics*. Harper and Bros., 170 pp.
- Bratseth, A., 1986: Statistical interpolation by means of successive corrections. *Tellus*, **38A**, 439–447.
- Charney, J., 1947: The dynamics of long waves in a baroclinic westerly current. *J. Meteor.*, **4**, 135–162.
- , R. Fjørtoft, and J. von Neumann, 1950: Numerical integration of the barotropic vorticity equation. *Tellus*, **2**, 237–254.
- Clebsch, A., 1857: Über ein allgemeine Transformation der Hydrodynamischen Gleichungen Crelle. *J. für Math.*, **54**, 293–312.
- Cressman, G., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- Demaria, M., 1996: A history of hurricane forecasting for the Atlantic Basin. *Historical Essays in Meteorology 1919–1995*, J. Fleming, Ed., Amer. Meteor. Soc., 263–305.
- Eckart, C., 1960a: *Hydrodynamics of the Oceans and Atmospheres*. Pergamon Press, 290 pp.
- , 1960b: Variation principles of hydrodynamics. *Phys. Fluids*, **3**, 421–427.
- Eddy, A., 1967: The statistical objective analysis of scalar data fields. *J. Appl. Meteor.*, **6**, 597–609.
- Eliassen, A., 1954: Provisional report on calculation of spatial covariance and autocorrelation of the pressure field: Appendix to Report No. 5. Videnskaps-Akademiets Institutt for Vaer-Og Klimaforskning, Oslo, Norway, 12 pp. [Available from Norwegian Meteorological Institute, P.O. Box 43, Blindern, N-0313 Oslo, Norway.]
- , 1982: Vilhelm Bjerknes and his students. *Annu. Rev. Fluid Mech.*, **14**, 1–11.
- , 1990: Oral history. Interviewer: J. Green. Royal Meteor. Soc., 43 pp. [Available from the Royal Society History Group, 104 Oxford Rd., Reading, Berkshire RG17LL, United Kingdom.]
- Evensen, G., 2007: *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag, 279 pp.
- Fisher, M., 2003: Background error covariance modeling. *Proc. Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean*, Reading, United Kingdom, ECMWF, 1–28. [Available from ECMWF, Shinfield Park, Reading RG29AX, United Kingdom.]
- Fjørtoft, R., 1952: On the numerical method of integrating the barotropic vorticity equation. *Tellus*, **4**, 179–194.
- Fujita, T., 1952: Pressure distribution in typhoon. *Geophys. Mag.*, **23**, 437–451.
- Gambo, K., 1950: On criteria for stability of the westerlies. *Geophysical Notes*, Vol. 3, Tokyo University, Tokyo, Japan, 29 pp.
- Gandin, L., 1965: *Objective Analysis of Meteorological Fields*. Israel Program for Scientific Translations, 242 pp.
- Gauss, C., 1809: *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium* (Theory of the Motion of Heavenly Bodies Moving about the Sun in Conic Sections). 326 pp. (English translation by C. H. Davis reissued by Dover Publications, 1963.)
- Ghil, M., S. Cohen, J. Tavantzis, K. Bube, and E. Isaacson, 1981: Application of estimation theory to numerical weather prediction. *Dynamic Meteorology: Data Assimilation Methods*, L. Bengtsson, M. Ghil, and E. Kallén, Eds., Springer-Verlag, 139–224.
- Gilchrist, B., and G. Cressman, 1954: An experiment in objective analysis. *Tellus*, **6**, 309–318.
- Haug, O., 1959: A method of numerical weather map analysis. Det Norske Meteorologiske Institutt Sci. Rep. 5, 10 pp.
- Hubert, L., 1959: An operational test of a numerical prediction method for hurricanes. *Mon. Wea. Rev.*, **87**, 222–230.
- Kalman, R., 1960: A new approach to linear filtering and prediction problems. *Trans. Amer. Soc. Mech. Eng.: J. Basic Eng.*, **82D**, 35–45.
- Kaplan, L., 1959: Influence of atmospheric structure from remote radiation measurements. *J. Opt. Soc. Amer.*, **49**, 1004–1007.
- Kasahara, A., 1957: The numerical prediction of hurricane movement with the barotropic model. *J. Meteor.*, **14**, 386–402.
- Kolmogorov, A., 1941: Interpolation and extrapolation. *Bull. Acad. Sci. USSR Ser. Math.*, **5**, 3–14.
- Lakshmvarahan, S., and D. Stensrud, 2008: Ensemble Kalman filter in meteorology: An overview. IEEE Control Systems Society on Kalman Filtering, Special Issue of the 50th Anniversary of Kalman Filtering, in press.
- Lamb, H., 1932: *Hydrodynamics*. 6th ed. Cambridge University Press, 738 pp. (Reprinted by Dover Publications, 1954.)
- Lanczos, C., 1970: *The Variational Principles of Mechanics*. University of Toronto Press, 418 pp.
- LeDimet, F., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110.
- Lewis, J., 1972: An operational upper air analysis using the variational method. *Tellus*, **24**, 514–530.
- , and J. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37A**, 307–322.
- , S. Lakshmvarahan, and S. Dhall, 2006: *Dynamic Data Assimilation: A Least Squares Approach*. Cambridge University Press, 654 pp.
- Lorenc, A., 1981: A global three-dimensional multivariate statistical analysis scheme. *Mon. Wea. Rev.*, **109**, 701–721.
- , 1986: Analysis method for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.
- , and F. Rawlins, 2005: Why does 4DVAR beat 3DVAR? *Quart. J. Roy. Meteor. Soc.*, **131**, 3247–3257.
- Lorenz, E., 1996: The evolution of dynamic meteorology. *Historical Essays on Meteorology 1919–1995*, J. Fleming, Ed., Amer. Meteor. Soc., 3–19.
- Lynch, J., 2006: *The Emergence of Numerical Weather Prediction (Richardson's Dream)*. Cambridge University Press, 279 pp.
- Panofsky, H., 1949: Objective weather map analysis. *J. Meteor.*, **6**, 386–392.
- Parrish, D., and J. Derber, 1992: The National Meteorological Center's spectral-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Phillips, N., 1956: The general circulation of the atmosphere: A numerical experiment. *Quart. J. Roy. Meteor. Soc.*, **82**, 123–164.
- Platzman, G., 1979: The ENIAC computations of 1950—Gateway to numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **60**, 302–312.
- Purser, R. J., 1984: A new approach to the optimal assimilation of meteorological data by iterative Bayesian analysis. Preprints,

- 10th Conf. on Weather Forecasting and Analysis*, Portland, OR, Amer. Meteor. Soc., 102–105.
- Rabier, F., 2005: Overview of global data assimilation developments in numerical weather prediction centers. *Quart. J. Roy. Meteor. Soc.*, **131**, 3215–3233.
- , and Z. Liu, 2003: Variational data assimilation: theory and overview. *Proc. Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean*. Reading, United Kingdom, ECMWF, 29–43. [Available from ECMWF, Shinfield Park, Reading RG29AX, United Kingdom.]
- , H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of 4D variational assimilation. Part I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143–1170.
- Richardson, L., 1922: *Weather Prediction by Numerical Process*. Cambridge University Press, 236 pp. (Reprinted by Dover Publications, 1965.)
- Rutherford, I., 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.*, **29**, 809–815.
- Sasaki, Y., 1955: A fundamental study of the numerical prediction based on the variational principle. *J. Meteor. Soc. Japan*, **33**, 30–43.
- , 1958: An objective analysis based on the variational method. *J. Meteor. Soc. Japan*, **36**, 1–12.
- , 1970a: Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.*, **98**, 875–883.
- , 1970b: Numerical variational analysis formulated under the constraints as determined by longwave equations and a low-pass filter. *Mon. Wea. Rev.*, **98**, 884–898.
- , 1970c: Numerical variational analysis with weak constraint and application to surface analysis of severe storm gust. *Mon. Wea. Rev.*, **98**, 899–910.
- , and K. Miyakoda, 1954: Prediction of typhoon tracks on the basis of numerical weather forecasting method. *Proc. Symp. on Typhoons*, Tokyo, Japan, UNESCO, 221–234.
- Schlatter, T., G. Branstator, and L. Thiel, 1976: Testing a global multivariate statistical objective analysis scheme with observed data. *Mon. Wea. Rev.*, **104**, 765–783.
- Schoppa, L., 1991: *Education Reform in Japan: A Case of Immoblist Politics*. Routledge Publishing, 319 pp.
- Smagorinsky, J., 1963: General circulation experiments with the primitive equations. I. The basic experiment. *Mon. Wea. Rev.*, **91**, 99–164.
- , 1971: Oral history. Interviewer: R. Mertz. National Museum of American History, Smithsonian Institution, Washington, DC, 20560, 100 pp. [Available from National Museum of American History, Smithsonian Institution, Washington, DC 20560.]
- , 1978: History and progress. The Global Weather Experiment—Perspectives on implementation and exploitation. FGGE Advisory Panel Rep., National Academy of Sciences, Washington, DC, 104 pp.
- Swirling, P., 1959: First-order error propagation in a stagewise smoothing procedure for satellite observations. *J. Astronaut. Sci.*, **6**, 46–52.
- Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.*, **113**, 1311–1328.
- Thacker, C., and R. Long, 1988: Fitting dynamics to data. *J. Geophys. Res.*, **93**, 1227–1240.
- Thompson, P., 1980: A short-range prediction scheme based on conservation principles and the generalized balance equation. *Contrib. Atmos. Phys.*, **53**, 256–263.
- Wahba, G., and J. Wendelberger, 1980: Some new mathematical methods for variational objective analysis using splines and cross validation. *Mon. Wea. Rev.*, **108**, 1122–1143.
- Wiener, N., 1947: *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Technology Press and John Wiley and Sons, 163 pp.
- Wiin-Nielsen, A., 1991: The birth of numerical weather prediction. *Tellus*, **43A**, 36–52.
- Yaglom, A., 1962: *An Introduction to the Theory of Stationary Random Functions* (translated from Russian by R. A. Silverman). Prentice-Hall, 235 pp.