Reconstructing 500-hPa Height Fields Over the Northern Hemisphere

Brian C. Polansky

A thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science

University of Washington

2002

Program Authorized to Offer Degree: Atmospheric Sciences

University of Washington

Graduate School

This is to certify that I have examined this copy of a master's thesis by

Brian C. Polansky

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

John M. Wallace

_____

James R. Holton

_____

Clifford F. Mass

Date: _____

In presenting this thesis in partial fulfillment of the requirements for a Master's degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Any other reproduction for any purpose or by any means shall not be allowed without my written permission.

Signature_____

Date_____

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my mom, step-mom, and brother for all their support throughout the years. It has come in many forms and all of them were appreciated to the fullest. I love you all.

It seems that taking my time to get my undergraduate degree proved to be a winning move. I was extremely fortunate to come into this program when I did. My happiest memories of the past two years will be of the people who began as my classmates, but have become much more. I am not sure if I would have made it without you.

Special thanks to Mike Wallace for taking me on as a student and helping me through this process in a timely manner. Also, a big debt of gratitude goes to Todd Mitchell who helped me make my way through the data. To everyone who shared these last two years with me, I thank you and wish you all the best of luck.

DEDICATION

To my father. I will always remember you watching the clouds.

Chapter 1

INTRODUCTION

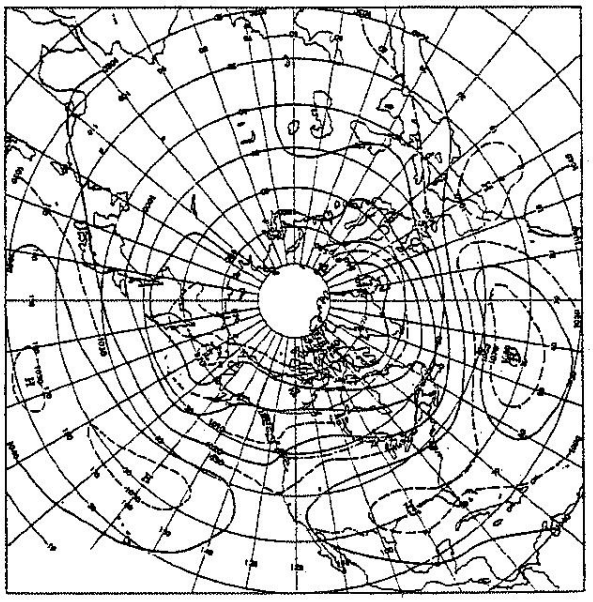

Figure 1.1: Reconstruction of 700-hPa height field for August 1936 by Namias.

The National Center for Atmospheric Research/National Centers for Environmental Prediction (NCAR/NCEP) reanalyzed 500-hPa height fields span the time period 1948-present. Attempts have been made to reconstruct upper level gph prior to that time interval based on surface data using various statistical and subjective methods.

A. Welch (U.K. Met. Office) generated 500-hPa geopotential height surfaces back to

the 1890s using surface air temperature (SAT), sea-level pressure (SLP) and the hypsometric equation to derive the height fields (pers. comm. J. Arnott). Namias reconstructed monthly 700-hPa gph back to 1930 using regression equations between SLP and 700-hPa gph. Figure 1 shows his reconstruction for the month of August 1936.

Reconstructions of daily 500-hPa height fields back to the late 18th century were attempted by Kington (1975), who inferred the 1000-500-hPa thickness from SAT and wet-bulb potential temperature. Derived thickness contours were advected using the 1000-hPa geostrophic wind. Further analysis of the time evolution of thickness patterns and their discontinuities were made by using the position of observed surface fronts. The derived thickness patterns were adjusted in order to make them consistent with the position and dynamic evolution of surface weather systems. Climatological data, based on monthly mean 500-hPa height, were used to adjust the reconstructions if extreme anomalies were produced.

A. Douglas used pilot balloon data (PIBAL) and observations of wind shifts from east to west to define upper level wind patterns and ridge positions back to the time period of the 1930s (Douglas pers. comm.). Currently, the PIBAL data are not accessible to researchers, but plans are under way to make them available. Douglas has suggested that the most accurate reconstructions would be achieved by using a combination of linear regression equations between SAT, SLP, and the upper level heights along with the PIBAL data.

Recent studies have reconstructed the gph field at various levels in the middle and upper troposphere based on multivariate linear regression. Klein and Dai (1998) attempted to reconstruct 700-hPa height fields over North America using SAT, SLP and precipitation from individual observation stations spaced across the domain. These variables were used in what was termed a ‘reverse specification’ in which the surface variables were allowed to specify the 700-hPa gph surface rather than the other way around, as was typical in operational weather prediction several decades ago. The inclusion of precipitation added very little to the accuracy of the reconstructions so these data were discarded. The 700-hPa height regressions were performed independently at each gridpoint. Heights at individual grid points were reconstructed using a regression equation unique to that grid point based

on SAT and SLP within the domain of the analysis. A step-wise procedure was used to maximize the explained variance while limiting the number of predictors used in the individual regression equations. The results were considered quite encouraging.

Schmutz et al (2001) attempted a similar reconstruction of the 300, 500, 700-hPa height fields over Europe and the Eastern North Atlantic based on SAT and precipitation from 25 selected stations along with gridded SLP data. Principal component analysis was performed to minimize the number of predictors while retaining maximal information about the surface fields. Canonical correlation analysis was performed to maximize correlations between the principal components of the surface and upper air data. This study was similar to Klein and Dai's, in that the predictors at the Earth's surface could influence the upper level gph fields, not only directly overhead, but anywhere in the domain of the analysis.

The purpose of this thesis is to explore the feasibility of reconstructing 500-hPa height fields in a straightforward manner by the use of surface data alone. Unlike previous studies, surface variables here are used only to reconstruct upper level height fields in the columns directly above them. It is of interest to determine the performance of this study compared to some of the previous studies mentioned above.

A list of the datasets and a description of methods used is described in section 2. The results of the models, including statistics, variance analysis, and cross-validation are presented in section 3. A reconstruction of a 500-hPa height field during the mid-1930s over North America is presented in section 4. Comparisons between the results of previous studies and this one are given in section 5. Conclusions are discussed in section 6.

Chapter 2

DATA AND METHODS

2.1 Data

The following datasets were used in this paper:

- NCEP reanalysis data for monthly averaged SLP, SAT, and 500 hPa gph anomalies for the years 1950-1995. The complete data represent the Northern Hemisphere from $20\,^{\circ}\mathrm{N}$ to the pole on a 144 x 29 latitude-longitude array with $2.5\,^{\circ}$ x $2.5\,^{\circ}$ resolution.

- Sea surface temperature (SST) and CRU05 land air temperature data for 1854-1997. Air temperature over land spanned the period 1901-95 and had an original latitude-longitude resolution of $0.5\,^{\circ}$. SST was obtained from the Comprehensive Ocean-Atmosphere dataset (COADS) which were originally on a $2\,^{\circ}$ x $2\,^{\circ}$ grid. The CRU05 data were obtained from the University of East Anglia. These data were combined over the domain into $2.5\,^{\circ}$ x $2.5\,^{\circ}$ resolution in order to be compatible with the NCEP data.

- Trenberth's Northern Hemisphere Monthly Sea-Level Pressure Grids from 1899 onward Obtained from http://dss.ucar.edu/datasets/ds010.1/. These data were originally on a $5\,^{\circ}$ x $5\,^{\circ}$ grid and were converted to $2.5\,^{\circ}$ x $2.5\,^{\circ}$ resolution.

2.2 Method

As in earlier approaches to this problem, linear multivariate regression is used to specify 500-hPa height, but in this case it is applied in a pointwise manner; i.e the regression is carried out independently at each grid point using data only for that grid point. All

calculations are based on anomaly fields. Hence, the seasonally varying climatology is built into the reconstructed 500-hPa height field.

Transformations of surface variables are based on results using the hypsometric equation:

$$Z_2 - Z_1 = \frac{R_d T_v}{g_0} ln \left(\frac{p_1}{p_2} \right) \tag{2.1}$$

where Z_2 and Z_1 are 500 and 1000-hPa levels respectively, p_2 and p_1 are pressure at the corresponding height levels, R_d is the gas constant for dry air, T_v is the virtual temperature, and g_0 is gravity.

Two different approaches are tried. The first regression model uses one variable, SAT, while the second one uses, SAT and SLP. In both cases the analysis is performed on monthly mean data.

For the regressions based on the NCEP data, the domain is defined as extending from $20\,^\circ$ to $90\,^\circ$N. First, monthly climatological mean values of all respective data sets are removed from each individual grid point to obtain anomaly fields for SAT, SLP and 500-hPa height. Climatological means of the datasets are based on the period of record 1950-1995, except for winter seasonal mean data which is based on 45 winter seasons 1950-51 through 1994-95. The regressions are performed on a point by point basis producing 4176 individual regression models for the NCEP data, one corresponding to each $2.5\,^\circ$ x $2.5\,^\circ$ grid point in space. They are calculated for the two calendar months of January and July and for the seasons of winter (DJF) and summer (JJA). The seasonal mean regression fields are formed simply by averaging the monthly mean regression fields over the winter (DJF) and summer (JJA) months.

The regressions and all the statistics presented here are then calculated using these anomaly fields. Maps showing correlation and regression coefficients along with root mean squared error are used to assess the accuracy of the models. Maps of anomaly and full fields are also shown for sample months.

For each grid point in space root mean squared error is calculated by:

$$RMSE = \left[\frac{\sum_{i=1}^{n}(y_i - y_i^*)^2}{n}\right]^{\frac{1}{2}} \tag{2.2}$$

where y is the predictand, y^* is the predicted height and n is number of months in the regression. Correlation coefficients for the individual grid points are calculated using:

$$\mathrm{corr}(y, y^*) = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(y_i^* - \overline{y}^*)}{\left[\sum_{i=1}^{n}(y_i - \overline{y})^2 \sum_{i=1}^{n}(y_i^* - \overline{y}^*)^2\right]^{1/2}} \tag{2.3}$$

with terms and indices defined as above.

2.3 The regression models

In order to explore the importance of SAT as a predictor of 1000-500-hPa thickness, a single variable, least-squares regression based on the model

$$Z_{500} - Z_{1000} = a_1(20.3T_s) + \epsilon_1 \tag{2.4}$$

or, equivalently

$$Z_{500} = Z_{1000} + a_1(20.3T_s) + \epsilon_1 \tag{2.5}$$

was performed, where Z_{500} and Z_{1000} are gph in units of meters at 500 and 1000-hPa respectively, a_1 is the regression coefficient, T_s is SAT in units $^\circ$C, and ϵ_1 is the error associated with the model which is to be minimized in a least-squares sense. The factor 20.3 associated with T_s is the numerical value of R_d/g_0 in (3.1).

Z_{1000} is calculated by scaling SLP, p_o, to meters using deviations from the 1000-hPa level. The hypsometric equation is again used to derive the approximation:

$$Z_{1000} \approx 8(p_o - 1000) \tag{2.6}$$

The two variable, linear regression model is based on the equation

$$Z_{500} = a_1(20.3T_s) + a_2 Z_{1000} + \epsilon_2 \tag{2.7}$$

or alternatively

$$Z_{500} = a_1^* \widehat{T}_s + a_2^* \hat{p}_o + \epsilon_2 \tag{2.8}$$

where $\widehat{T}_s$ and $\hat{p}_o$ are standardized to unit variance and

$$a_1^* = a_1[std(20.3T_s)] \tag{2.9}$$

$$a_2^* = a_2[std(p_o)] \tag{2.10}$$

are coefficients of the standardized predictors. The regressions are performed on time series for individual grid points, just as in the one variable case.

The procedure is repeated for the two variable model using predictor datasets independent of the NCEP/NCAR reanalysis, consisting of SLP and SAT from the datasets described in section 2.1. The NCEP dataset was again used as the predictands. The regressions are performed on a grid covering the Northern Hemisphere from $70\,^\circ$ to $20\,^\circ$N. For grid points poleward of 70°N, observations are not considered reliable enough to be used as a basis for the regressions (Trenberth and Paolino 1980).

Finally, to test the sensitivity of the two variable model, the results of the model are subjected to cross-validation. The month for which the 500-hPa height field is to be specified is removed from the dataset prior to performing the regression, and the resulting regression coefficients were used in the specification. Root mean squared error and correlation values are calculated for the new cross-validated set and compared with their counterparts based on the complete data. The one predictor, two predictor and cross-validated fields are then compared.

Chapter 3

RESULTS

3.1 The one variable model

The single variable model results are presented here. In the model, SAT is being used as a proxy for the mean virtual temperature of the 1000-500-hPa layer. As was mentioned earlier, the regression was performed with SAT scaled by a factor of 20.3 to make its relationship to thickness more explicit. Therefore, if SAT were representative of the mean virtual temperature of the layer we should expect to obtain regression coefficient values near unity. The scaling also renders the regression coefficients dimensionless. Regression coefficients are shown in Fig. 3.1.

During January, values greater than one are found over most of the oceans, where the thermal inertia acts in a manner to damp SAT anomalies relative to the anomalies in the

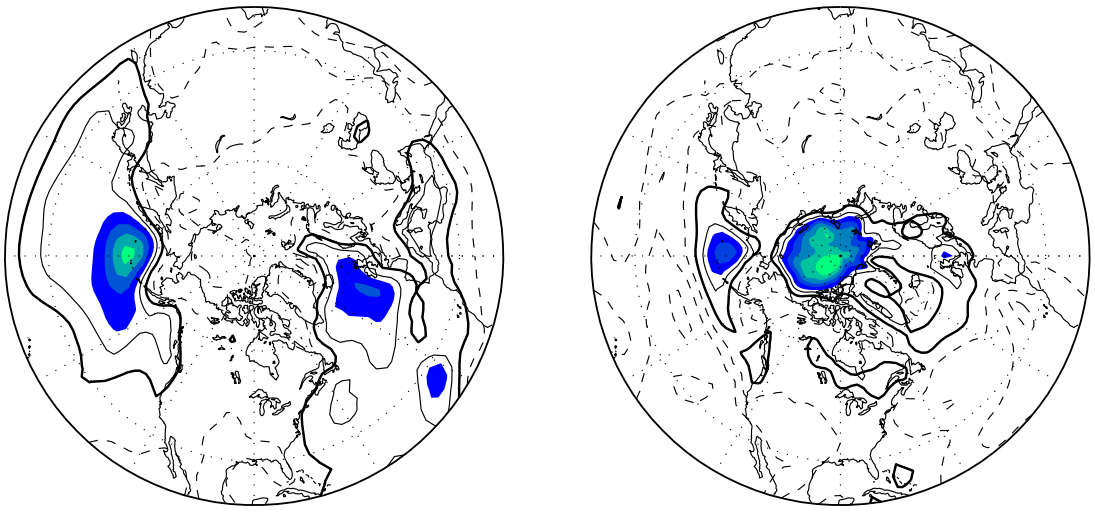

Figure 3.1: Regression coefficients for 1000-500-hPa thickness upon surface air temperature for January (left) and July (right), scaled as described in text. Heavy solid contour indicates a value of one. Dashed contours indicate values less than one. Contour interval 0.3.

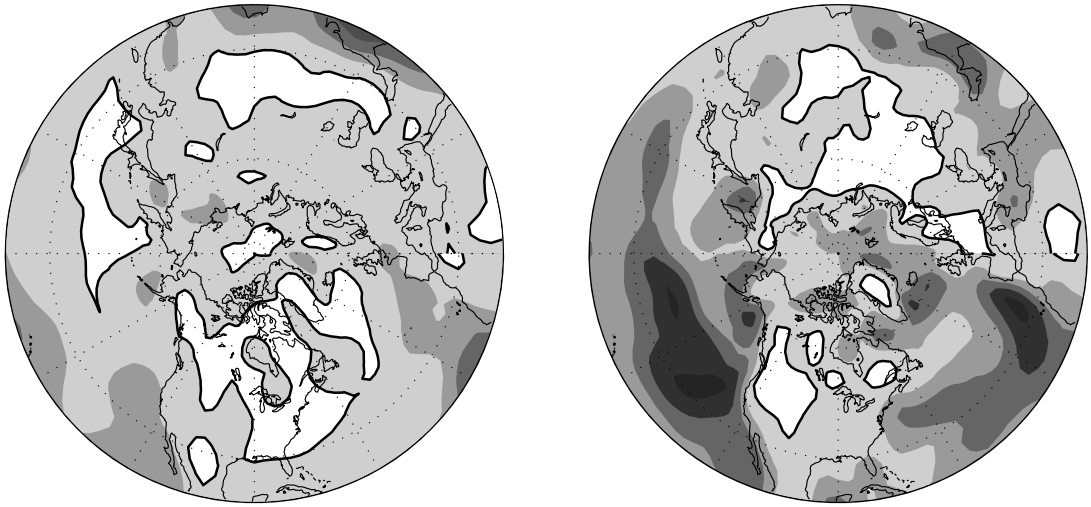

Figure 3.2: Correlation between analyzed and reconstructed 1000-500-hPa thickness anomalies based on one variable model (1000-500-hPa thickness regressed on surface air temperature) for January (left) and July (right). Unshaded areas indicate areas of correlation greater than 0.9. Heavy solid contour indicates value of 0.9. Contour interval 0.2

overlying layer. The model must therefore yield coefficients > 1.0 in order to capture the large deviations of the 1000-500-hPa thickness. A positive extremum with values in excess of 2.0 is observed in an area coinciding with the Aleutian Low and another positive extremum with values of ~ 1.8 is observed coinciding with the position of the Icelandic Low. Values near unity are found along coast lines, and over the continents values range from 0.6 to 1.0.

For July, the lobe corresponding to the Aleutian low is again present but extends over a smaller area with a positive extremum ~ 1.8. A very strong positive extremum, with values in excess of 2.6, is located over the Arctic. The differentiation between water and land is not as evident during July. Values less than unity are observed not only over most land areas, but also over the oceans equatorward of $50\,^{\circ}\mathrm{N}$. Values ranging as low 0.1 are observed off the coasts of California and North Africa. The low values in those regions reflect the weak linear correlation between SAT and upper level gph.

Correlation maps between observed and predicted 1000-500-hPa thickness anomalies for the months of January and July using this regression model with SAT as the predictor are shown in Fig. 3.2.

For January, SAT is correlated at a level in excess of 0.7 with thickness over most of the domain with the lowest values of 0.5 in the sub-tropics. Highest correlations are found primarily over land and over the North Pacific. The areal weighted average of correlation over the domain poleward of $30\,^{\circ}\mathrm{N}$ is 0.85 for the month of January (see table 3.1), indicating that temperature alone therefore explains approximately over 70% of the variance of 1000-500-hpa thickness.

During July, correlation coefficients of 0.5 and lower are observed over the eastern portions of the Northern Pacific and Atlantic oceans, with the lowest values of less than 0.2 directly off the west coasts of Africa and North America. These locations experience persistent layers of stratus clouds throughout the summer months, a result of high static stability which is generally found here in that season. The model was constrained using only SAT as an estimate of the total thickness of the layer. The presence of these clouds act to decouple the surface and upper atmosphere, which most likely led to the relatively poor performance of the model in these areas. The highest correlations for the month of July are observed over land with values in excess of 0.8 and 0.9 in many areas. Averaged over the domain poleward of $30\,^{\circ}\mathrm{N}$ the correlation was 0.71 for July (see table 3.1) indicating that roughly 50% of the variance is explained by the model.

The reconstructed 1000-500-hPa thickness fields were visually inspected in order to assess their fidelity. Areal averaged spatial correlations were calculated between these observed and reconstructed fields. Averaged over the entire dataset the correlation for January was 0.84; with values ranging from 0.95 for January 1951 to as low as 0.66 for January 1965. For July the mean is 0.77 and values for individual months range from 0.83 in 1954 to 0.67 in 1982. As a representation of the performance of the single variable model, sample fields for the months of January 1959 and July 1968 are displayed. These months were chosen for display because their spatial correlations fall very close to the mean for their respective calendar months (0.85 and 0.78 respectively). The anomaly fields for these months are shown in Figs. 3.3 and 3.4. The observed and reconstructed patterns for January 1968 are marked by positive thickness anomalies over the Arctic flanked by two areas of negative anomalies over North America and Northern Eurasia. The reconstructed field exhibits

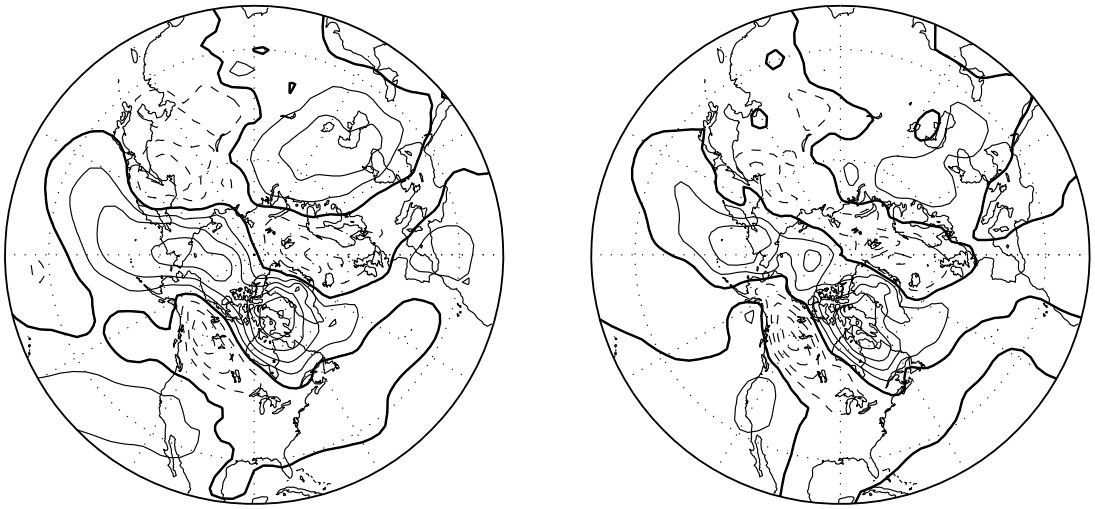

Figure 3.3: Analyzed 1000-500-hPa thickness anomalies for January 1959 (left) and reconstructed thickness anomalies for same month (right) based on single variable model (1000-500-hPa thickness regressed on surface air temperature). Heavy solid contour indicates a value of zero. Dashed contour indicates values less than zero. Contour interval 30 meters.

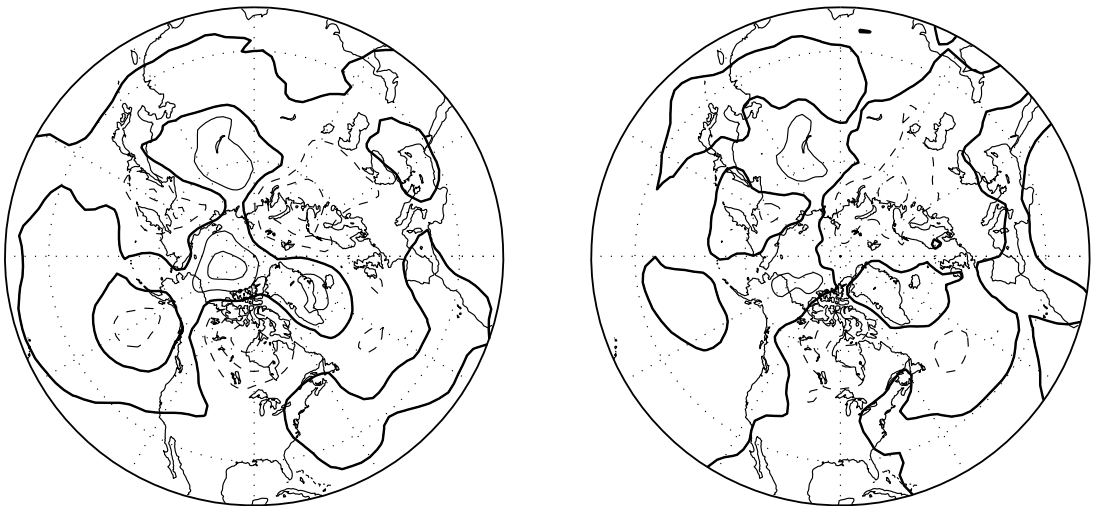

Figure 3.4: Same as Fig. 3.3 except for July 1968.

larger departures from the analyzed field towards lower latitudes. The largest difference, in excess of 60 meters, is found over southeastern Europe.

For July 1968, the model reconstruction also represents the major features in the analyzed field. The largest discrepancies are over the Arctic Ocean and over the North Pacific. SAT evidently provides a substantial amount of useful information.

The corresponding differences between the analyzed and reconstructed 1000-500-hPa thickness are shown in Fig. 3.5. These months can be considered representative of the errors between the reconstructed fields data produced by the one variable model and the NCEP reanalysis. Now we consider the reconstructions based on a regression model in which coefficients for both SAT and SLP are calculated at each grid point.

3.2 The multi-variable model

Figs. 3.6 and 3.7 show maps of regression coefficients for the months of January and July respectively for the two variable model. The temperature patterns are remarkably similar to their counterparts in the one variable model, shown in Fig. 3.1. In January, SLP

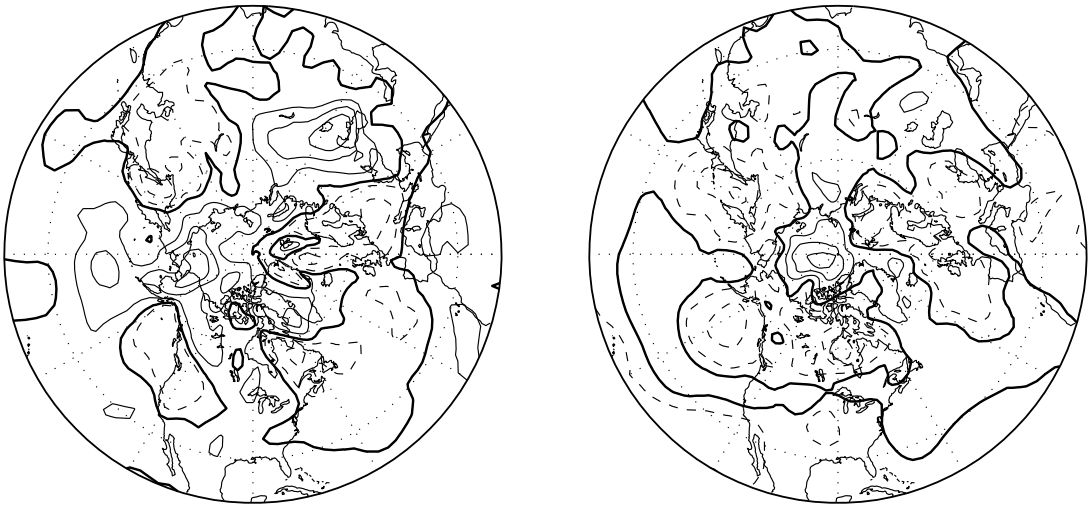

Figure 3.5: Difference between analyzed and reconstructed 1000-500-hPa thickness based on single variable model (1000-500-hPa thickness regressed on surface air temperature). January 1959 (left) and July 1968 (right). Heavy solid contour indicates a value of zero. Dashed contours indicate values less than zero. Contour interval 30 meters.

coefficients are greater than 1.0 over most of the hemisphere, indicating a tendency for fluctuations in the gph field to amplify with height. During July, SLP coefficients remain above unity in high latitudes but they are near or slightly below unity over most of the lower latitudes, with locally lower values.

In order to compare the relative importance of the two predictors, the ratio of the standardized SAT to SLP coefficients are shown in Fig. 3.8. Both coefficients are non-dimensionalized by use of relations derived from the hypsometric equation as stated earlier. Values less than unity are indicative of more barotropic conditions and values greater than one correspond to more baroclinic conditions.

During January, more baroclinic conditions are observed over land, and especially to the leeward side of major mountain ranges. More barotropic conditions are observed over most of the oceans. This pattern is remarkably similar to that for the correlation coefficient between 1000 and 500-hPa heights shown in Fig. 3.9. (see also Fig. 2 of Blackmon et al. (1979)). In July, coefficients less than unity cover most land areas except the Asian monsoon region. More equivalent barotropic conditions are located over the northeastern areas of the Pacific and Atlantic oceans.

Correlations between the reconstructed 500-hPa heights of the two predictor model and analyzed data, shown in Fig. 3.10, reveal the good performance of this model. During January, the lowest values (~ 0.7) are found off the west coast of Mexico and the northwest coast of Africa. Over a large area of the Northern Hemisphere values are in excess of 0.9 with most of the North Atlantic and Pacific exhibiting values greater than 0.95.

In the month of July, the performance is not as good, but is substantially better than that for the one predictor model. The lowest values are located off the west coasts of North America and Africa. Correlation maps for summer (JJA) and winter (DJF) (not shown) are similar to those for July and January (see table 3.1).

Examples of reconstructed monthly mean anomaly fields are shown in Fig. 3.11 and reconstructed total height fields in Fig. 3.12 for January 1959 and in Figs. 3.13 and 3.14 for July 1968. Spatial correlations were calculated between the analyzed and reconstructed anomaly fields from the two variable model in the same manner as described in the previous

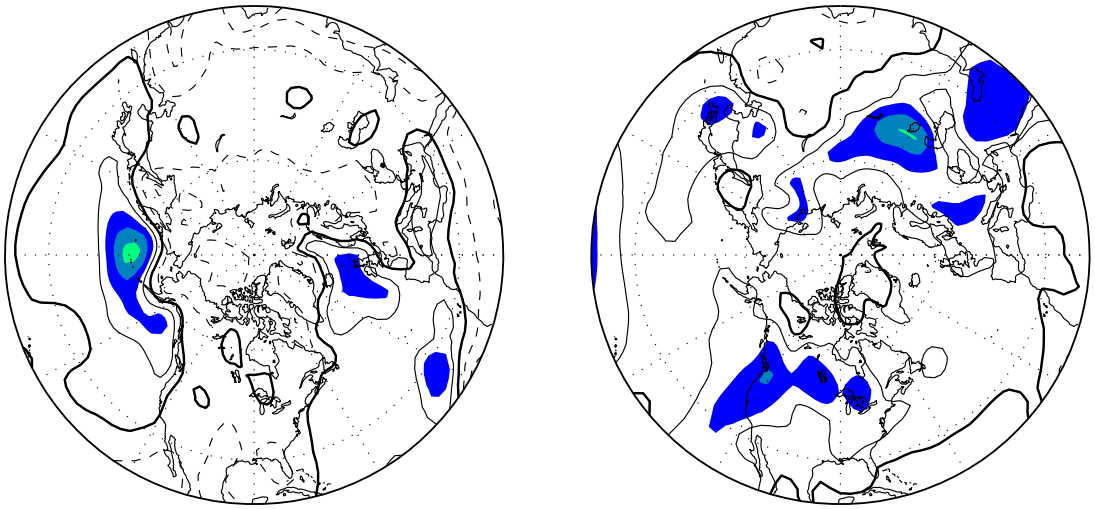

Figure 3.6: Regression coefficients for January. Surface air temperature (left) and sea-level pressure (right). Heavy solid line indicates a value of 1. Dashed lines indicates values less than 1. Contour interval 0.3

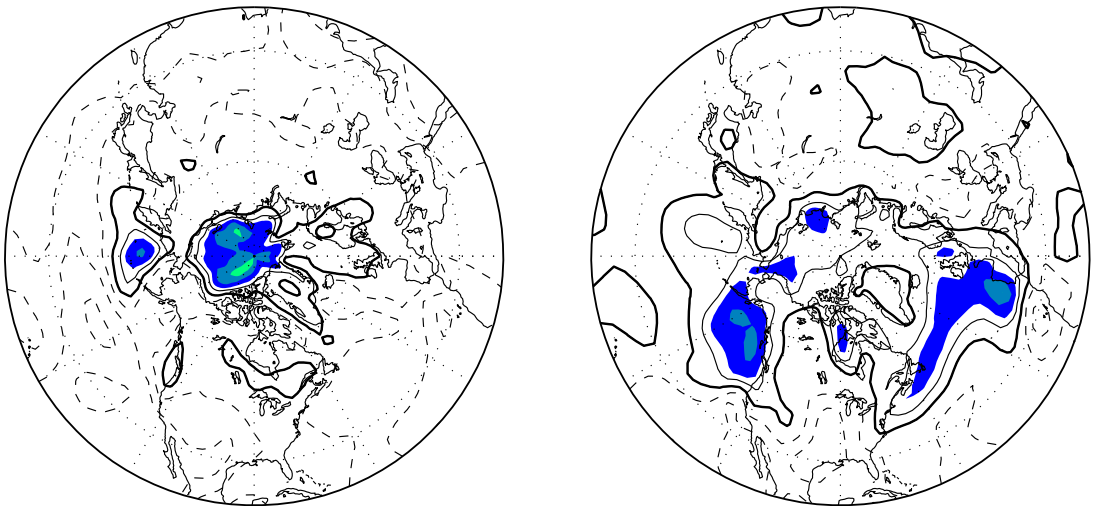

Figure 3.7: Same as Fig. 3.6 except for July.

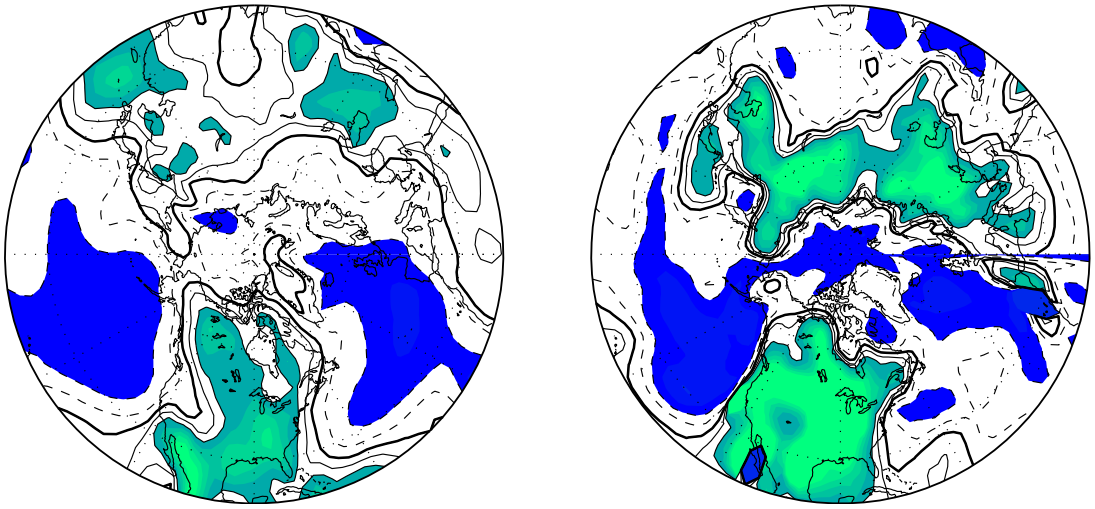

Figure 3.8: Ratio of standardized, nondimensional surface air temperature to sea-level pressure regression coefficients for January (left) and July (right). Heavy solid line indicates a value of one. Dashed contours indicate values less than one. Contour interval 0.25

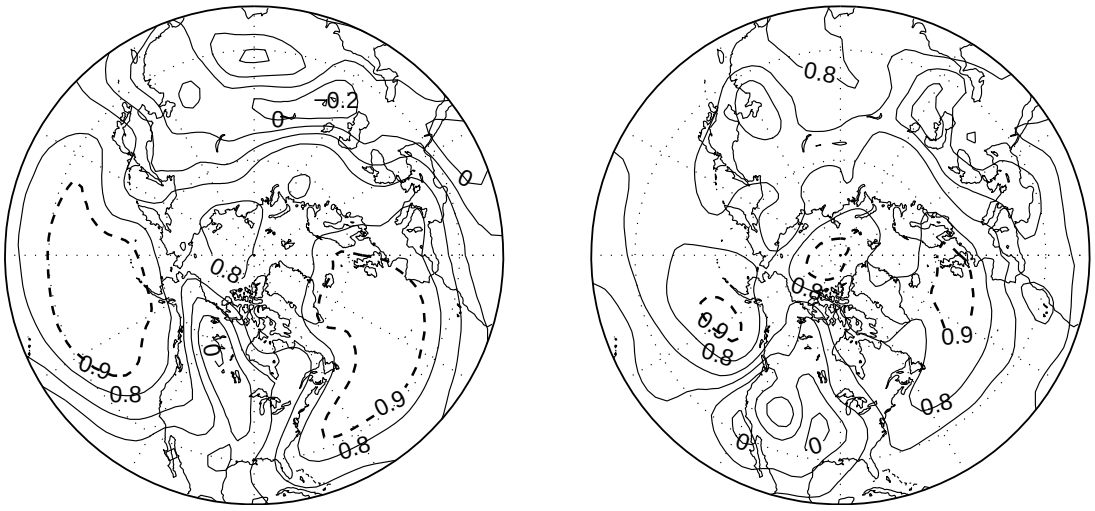

Figure 3.9: Correlation coefficient between 1000 and 500-hPa heights based on monthly mean data for winter (left) and summer (right). Contour interval 0.2.

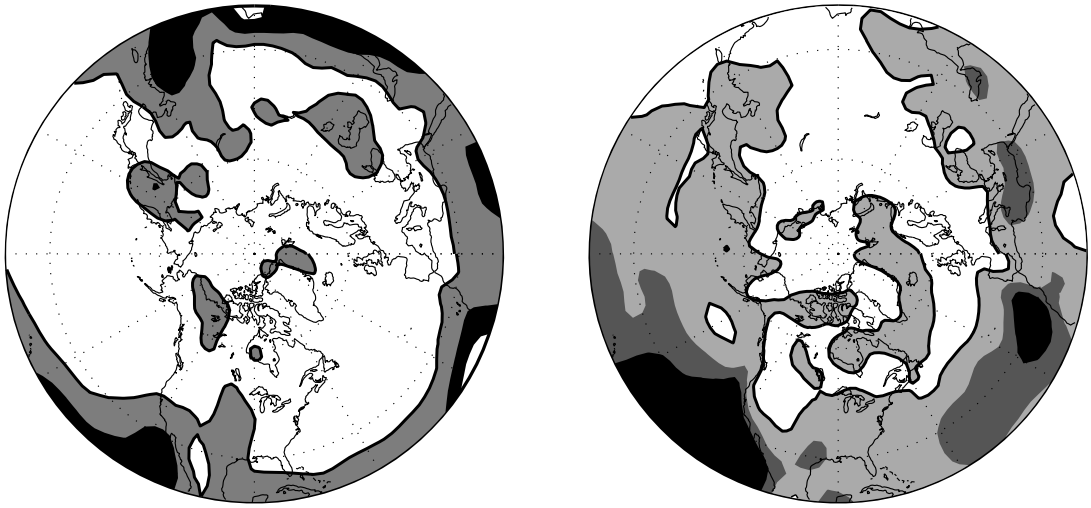

Figure 3.10: Correlation between analyzed and reconstructed 500-hPa height anomalies using two variable model (500-hPa heights regressed on surface air temperature and sea-level pressure) for January (left) and July (right). Heavy solid line indicates value of 0.9. Contour interval 0.2

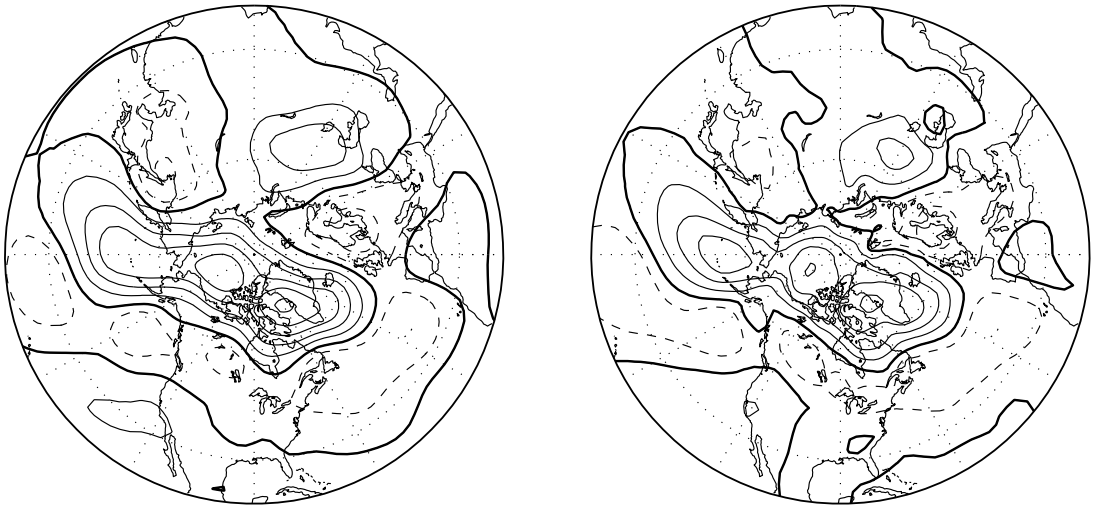

Figure 3.11: Analyzed (left) and reconstructed (right) 500-hPa height anomalies for January 1959 based on two variable model (500-hPa height regressed on surface air temperature and sea-level pressure). Heavy solid line indicates a value of zero. Dashed contours indicate negative values. Contour interval 40 meters.

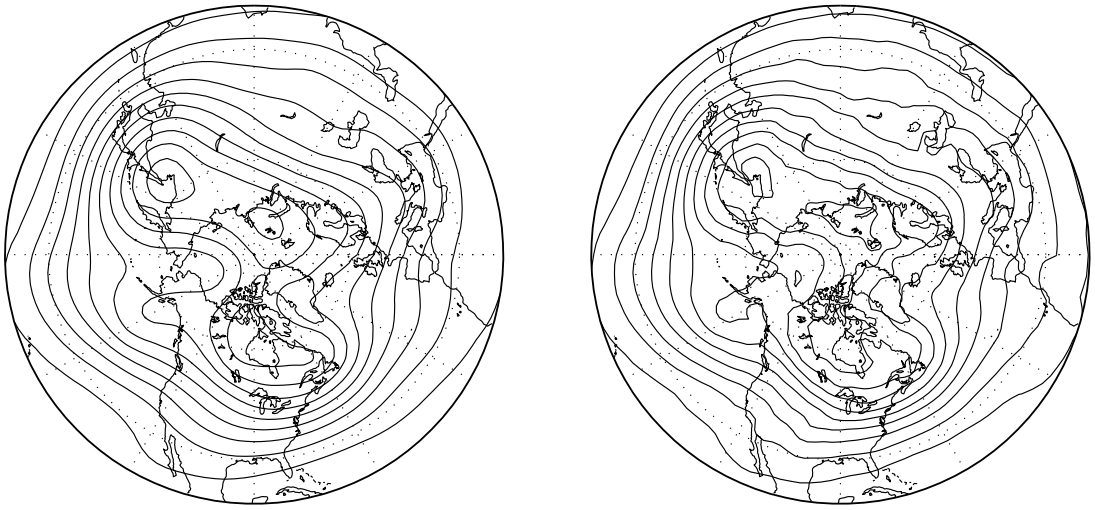

Figure 3.12: Analyzed (left) and reconstructed (right) 500-hPa heights for January 1959 based on two variable model (500-hPa height regressed on surface air temperature and sea-level pressure). Contour interval 80 meters.

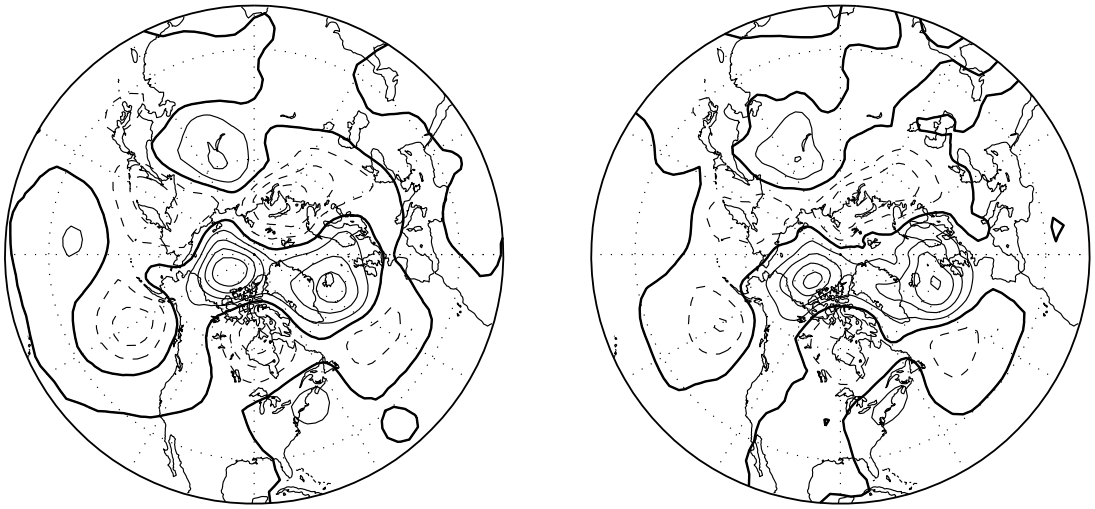

Figure 3.13: Same as Fig. 3.11 except for July 1968. Contour interval 30 meters.

section. In January values ranged from 0.84 for January 1965 to 0.97 for January 1950. The corresponding range for July was from 0.79 (1982) to 0.97 (1960). The mean for January was 0.94 and for July was 0.89, a substantial improvement from the one variable model. The values for months shown here are 0.95 for January 1959 and 0.89 for July 1968. The value of adding a second predictor, SLP, is clearly evident.

For both sample months there is strong agreement between the major features in the observed and reconstructed fields. The high heights that prevailed over the Arctic region in January 1959 are well captured by the two variable model as well as much of the structure over lower latitudes. The July features are also well represented, especially in the high latitudes. As borne out by the statistics and maps of other months, structures are best captured towards higher latitudes where correlation coefficients are higher. Error fields for the sample months are shown in Fig. 3.15. For July 1968 the largest errors, found just west of the British Isles, involve an eastward displacement of a ridge in the reconstruction. Again, as in the one variable model, the areas of largest error correspond to the areas of high variability in the higher latitudes. On average, these errors were smaller for the two predictor model.

3.3 The multi-variable model with independent data

The built in dynamical consistency between the 500-hPa height, SLP and SAT fields in the NCEP reanalysis contributes to the good performance of the reconstructions described in the previous sections. In order to make a more realistic assessment of the quality of reconstructed fields, we repeated the analysis using SLP from Trenberth and SAT from the combined datasets of COADS and CRU05. The regression analysis based on these data sets is implemented on a domain over the Northern Hemisphere from 20° to 70°N, which corresponds to a 21 x 144 latitude/longitude grid. The regression was performed as it was in the previous two predictor case. The region north of 70°N was excluded because the data coverage there is incomplete.

Results are presented along with the other model results in table 3.1. A modest increase in root mean squared error is noted in both months along with decreases in correlation

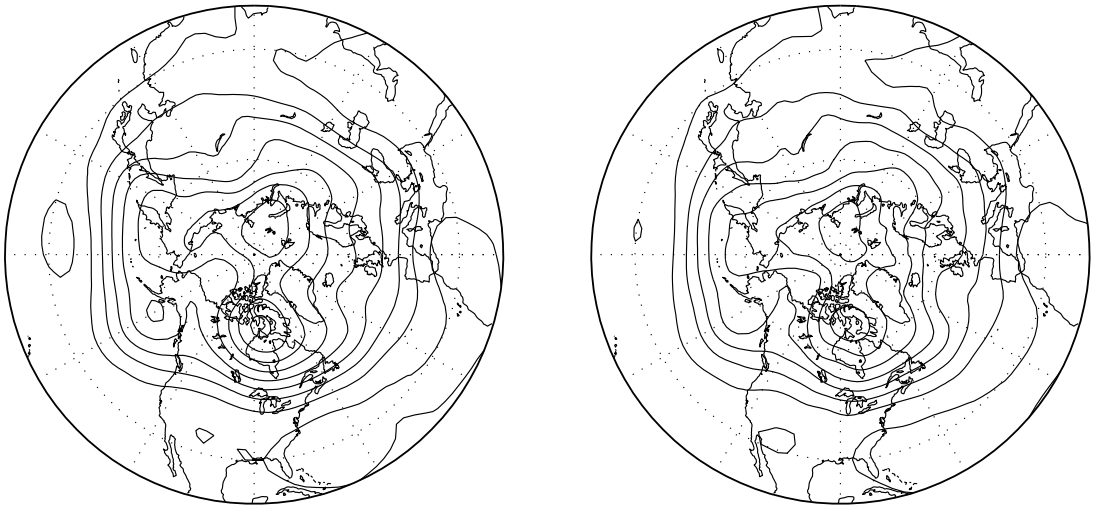

Figure 3.14: Same as Fig. 3.12 except for July 1968. Contour interval 60 meters.

Figure 3.15: Difference between analyzed and reconstructed anomalies based on two variable model (500-hPa height regressed on surface air temperature and sea-level pressure) for January 1959 (left) and July 1968 (right). Heavy solid line indicates a value of zero. Dashed contours indicated values less than 0. Contour interval 20 meters.

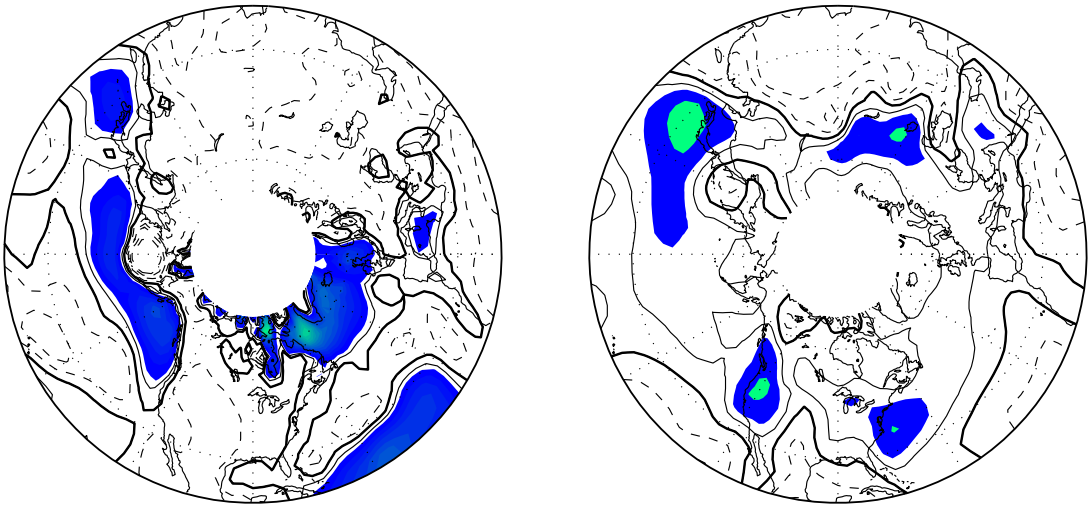

Figure 3.16: January regression coefficients from 500-hPa heights regressed on Trenberth (SLP) and COADS/CRU05 (SAT) data associated with surface air temperature (left) and sea-level pressure (right). Heavy solid line indicates a value of 1. Dashed lines indicates values less than 1. Contour interval 0.3

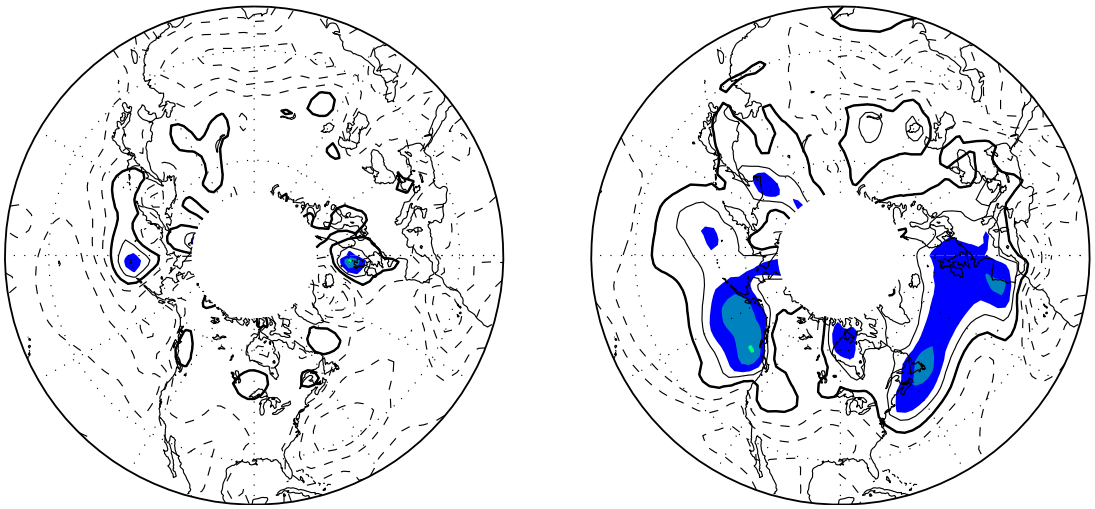

Figure 3.17: Same as Fig. 3.16 except for July.

and R^2 values compared to the corresponding statistics for the model based on SLP and SAT fields derived from the NCEP/NCAR reanalysis. Plots of the regression coefficients for January and July, shown in Figs. 3.16 and 3.17 respectively, can be compared with corresponding plots based on reanalysis data shown in Figs. 3.6 and 3.7.

In January, highest values of SAT coefficients are observed over areas corresponding to extrema based on the model using reanalysis data. Both have largest values located over water. The largest discrepancies are located over the mid-latitude oceans where values less than one are observed for the independent data. SAT coefficients are smaller on average for the independent data than those based on the reanalysis data. The overall pattern for SLP coefficients is similar between the two. Again, as in the SAT coefficient map, the values are smaller overall based on the independent data. The largest differences are generally located over land. In July, the SAT and SLP coefficient patterns for the two models are similar in both structure and magnitude.

3.4 Error statistics

Maps showing root mean squared error results from the two variable model based on SAT and SLP fields from the reanalysis are shown in Fig. 3.18. For January, root mean squared error ranges up to approximately 45 meters, with the higher values over land. For July, the errors range up to 30 meters over the higher latitudes (see table 3.1).

3.5 Cross validation

A cross-validation procedure was performed as described in section 2.2 to determine the robustness of the two variable model. Correlations between reconstucted and observed 500-hPa heights and root mean squared error values of the cross-validation dataset were averaged over the Northern Hemisphere poleward of $30\,^{\circ}\mathrm{N}$ and compared to the one and two variable model (table 3.1). This domain excluded much of the relatively low values of root mean squared error found in the lower latitudes.

Only small increases in root mean squared error are observed in the cross-validated set,

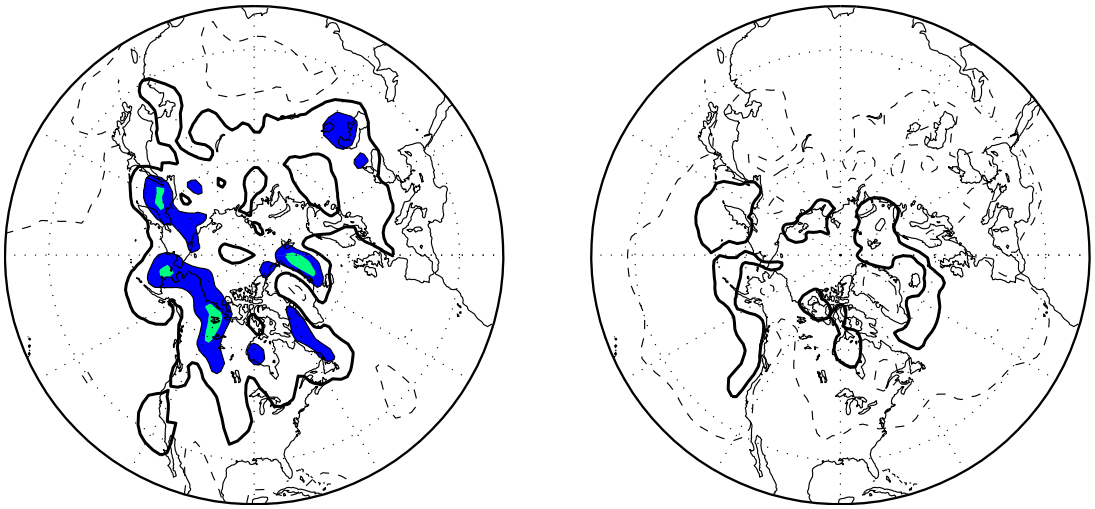

Figure 3.18: Root mean squared error for two variable model (500 millibar heights regressed on surface air temperature and sea-level pressure). January (left) and July (right). Heavy solid line indicates value of 20 meters. Dashed lines indicate values less than 20 meters. Contour interval 10 meters.

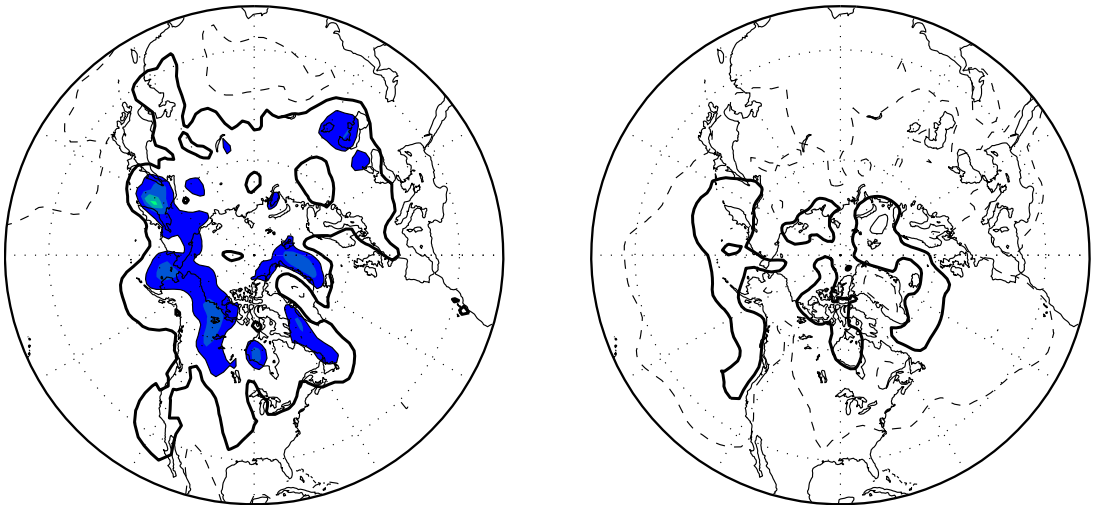

Figure 3.19: Cross-validated root mean squared error based on two variable model (500-hPa height regressed on surface air temperature and sea-level pressure). January (left) and July (right). Heavy solid line indicates value of 20 meters. Dashed lines indicate values less than 20 meters. Contour interval 10 meters.

adding confidence in the robustness of the two variable model. Shown in Fig. 3.19 is the root mean squared error map for this dataset.

Table 3.1: Areal average of correlation, root mean squared error, and R^2 values for the domain of the Northern Hemisphere poleward of $30\,^{\circ}\mathrm{N}$ for one, two, and cross-validated models using NCEP reanalysis data. Values for model using Trenberth (SLP) and COADS/CRU05 (SAT) are averaged over the Northern Hemisphere from $70\,^{\circ}$ to $20\,^{\circ}\mathrm{N}$.

Model	**Month(s)**	**Correlation**	**RMSE**(meters)	R^2 (%)
One Predictor	January	0.85	22	73
	July	0.71	14	54
	DJF	0.85	17	73
	JJA	0.74	10	57
Two Predictor	January	0.94	19	88
	July	0.86	13	76
	DJF	0.93	13	88
	JJA	0.87	9.3	76
Cross-Validation	January	0.93	20	87
Two Predictor	July	0.85	14	75
SLP(Trenberth)	January	0.82	24	67
SAT(COADS/CRU05)	July	0.69	15	51

3.6 Analysis of Variance

Shown in Fig. 3.20 is the variance of the 500-hPa height field of the reanalysis data for January and July. Figs. 3.21 and 3.22 show the variance patterns for the cross-validated reconstructed height based on the reanalysis and independent data respectively. The variance patterns of the observed heights are reproduced to a relatively high degree of accuracy in both models based on the reanalysis and independent data. The amplitude of extrema in the reconstructed heights based on the independent data are slightly smaller overall.

Empirical orthogonal function (EOF) analysis was performed on the winter (DJF) anomalies of the analyzed and cross-validated reconstructed heights based on the reanalysis data. The first two EOFs of each dataset are shown in Fig. 3.23. The leading EOF patterns are

similar in structure and amplitude. Shown in table 3.2 is the percent of variance explained by the first two EOFs of both datasets. The percentage of variance explained by the two fields are approximately equal for the leading EOFs.

EOF analysis was performed on the error field between the analyzed and reconstructed 500-hPa gph fields for winter (DJF). The resulting first two EOFs explain a relatively small fraction of the variance. EOF1 of the error field is shown in Fig. 3.24. This EOF contains little coherent structure across the Northern Hemisphere.

Correlation values were calculated between the corresponding PCs of the two datasets.

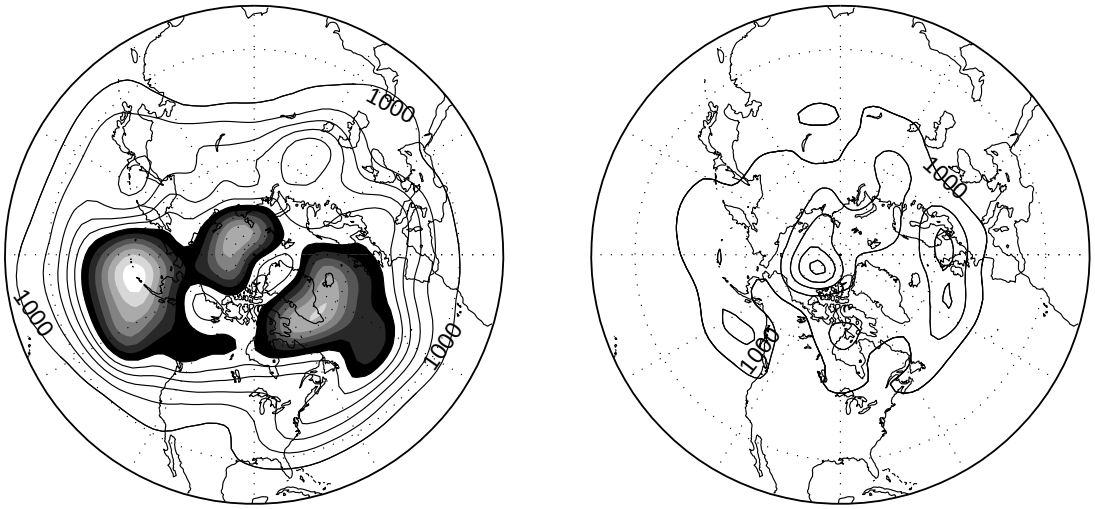

Figure 3.20: Variance of analyzed 500-hPa height for January (left) and July (right). Shaded values are greater than 6000. Contour interval 1000. Units are in m^2.

Table 3.2: Explained variance by first two EOFs of reconstructed 500-hPa geopotential height based on two predictor model using NCEP data, analyzed 500-hPa geopotential height, and error between the two fields.

Height field	EOF	Variance Explained (%)
Analyzed	1	18.4
	2	12.4
Reconstructed Multivariate	1	19.2
	2	12.3
Difference of two	1	8.6
	2	7.0

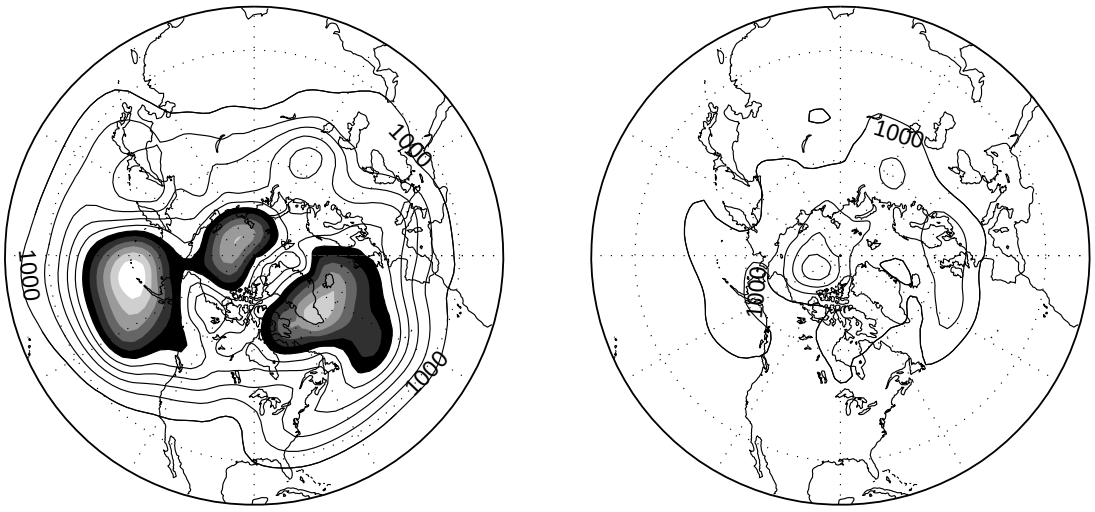

Figure 3.21: Same as figure 3.20 except for cross-validated reconstructed 500-hPa heights based on reanalysis data.

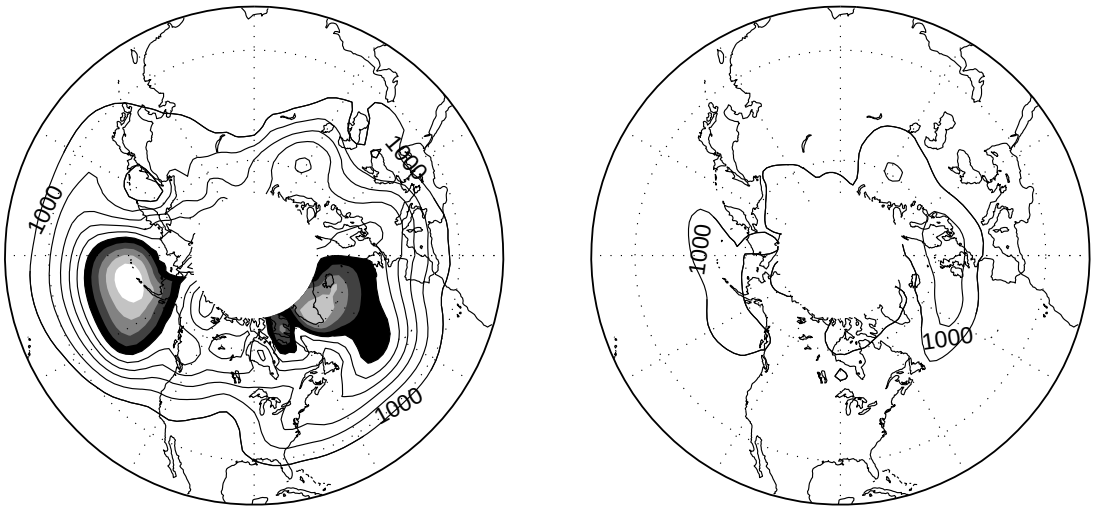

Figure 3.22: Same as figure 3.20 except for reconstructed 500-hPa heights based on independent data.

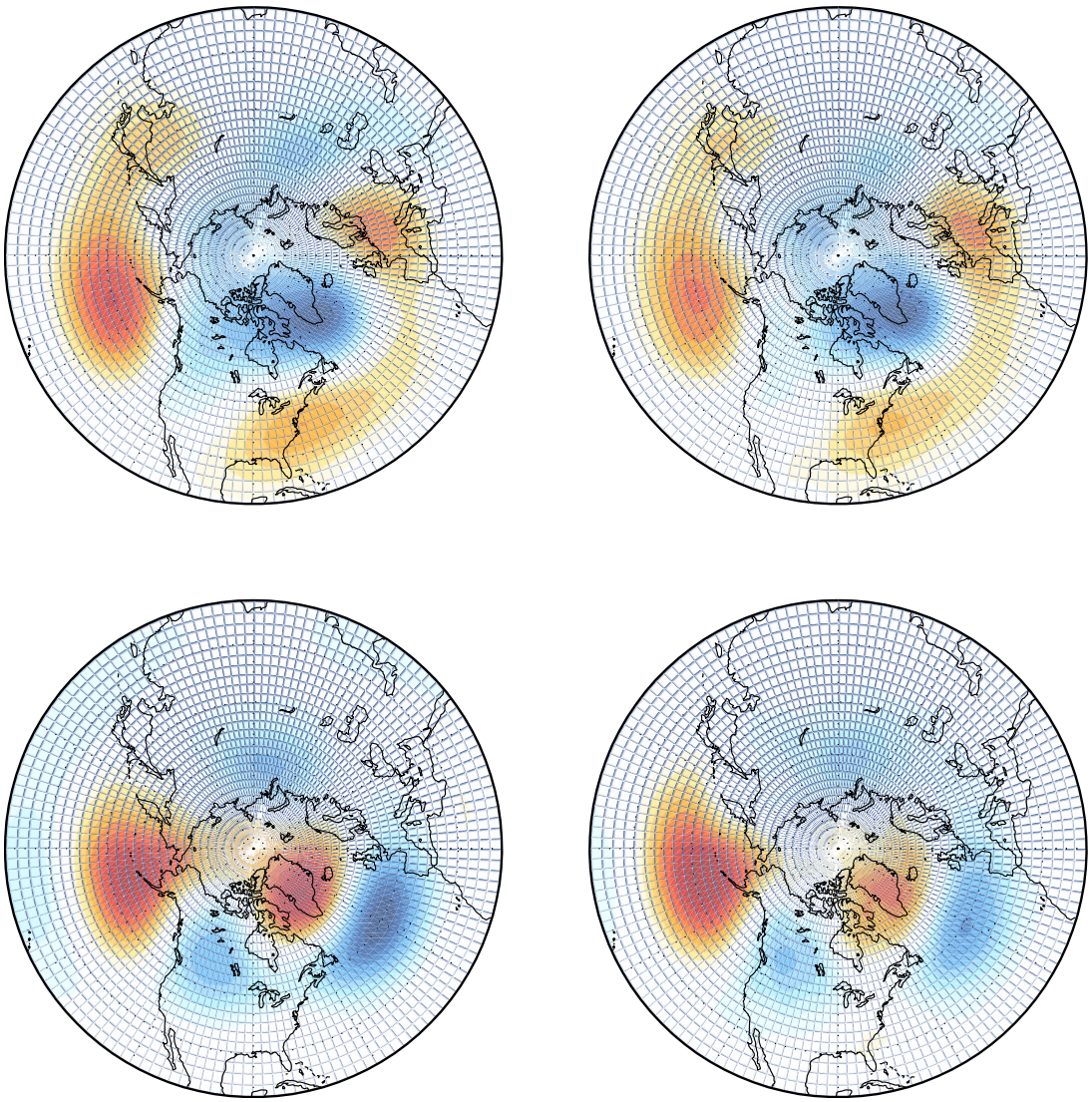

Figure 3.23: Top left, EOF1 of analyzed 500-hPa height anomalies. Top right, EOF1 of reconstructed height anomalies based on two variable model. Bottom left, EOF2 of analyzed 500 hPa geopotential height anomalies. Bottom right, EOF1 of reconstructed height anomalies.

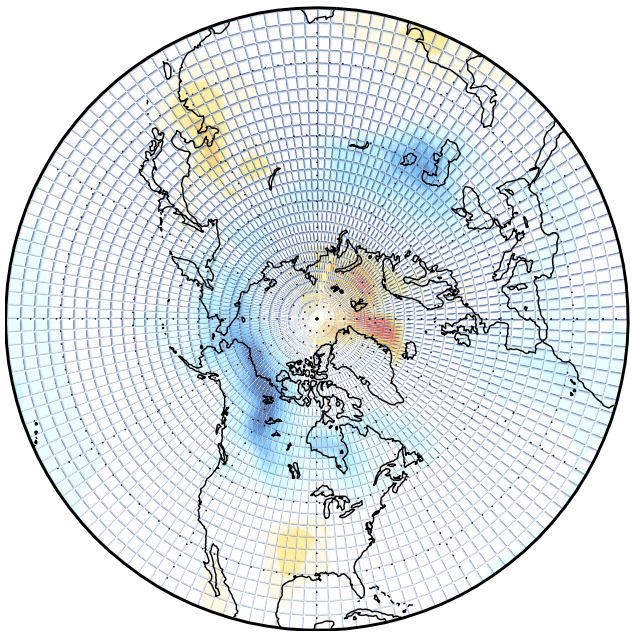

Figure 3.24: Leading EOF of error field between analyzed and reconstructed cross-validated 500-hPa geopotential height anomalies.

The correlation between PC1 of the analyzed and reconstructed datasets is 0.98 and for PC2 the value is 0.93. To determine whether the reconstructed PCs of 500-hPa height provide any useful information beyond what is contained in the SLP field alone, we performed multiple regression of the observed 500-hPa height PCs using PC1 and PC2 of SLP as predictors. PC1 of 500-hPa height, as reconstructed from PC1 and PC2 of SLP in this manner, was correlated with the observed PC1 at a level of 0.95. The corresponding result of PC2 was 0.84. Hence, the 500-hPa height PCs, as reconstructed by the method described in this thesis, are superior to PCs reconstructed from the SLP PCs.

Chapter 4

RECONSTRUCTION OF 500-HPA HEIGHT FIELDS FOR THE 1930S

4.1 Conditions present at the surface

The motivation for this study was to explore the possibility of extending upper level gph fields backward in time beyond the time period for which they are currently available. Shown here is the reconstruction for the period coinciding with the with the warmest three summers associated with the severe drought experienced over the Great Plains in the mid-1930s. Applying the results of the model to periods prior to 1950, this study, as in previous ones, presumes the relationship between the surface variables and 500-hPa gph has remained static over time.

Plotted in Figs. 4.1 and 4.2 are total and anomaly SAT and SLP fields over the North American sector for the nine summer months June, July, and August 1934-1936 based on COADS/CRU05 and Trenberth data. During these months, an area of warm temperature anomalies was located directly over the Great Plains of the North America with anomalies up to $\sim 2.5\,^{\circ}\mathrm{C}$.

4.2 500-hPa reconstruction

The reconstruction was performed in the same manner as the earlier seasonal data. The regression coefficients used are those from the results of the two variable model based on Trenberth SLP and COADS/CRU05 SAT for the time period 1950-1995.

Shown in Fig. 4.3 is the reconstructed 500-hPa anomaly and total gph field. A ridge of high pressure is observed over the United States with a maximum located over Northern

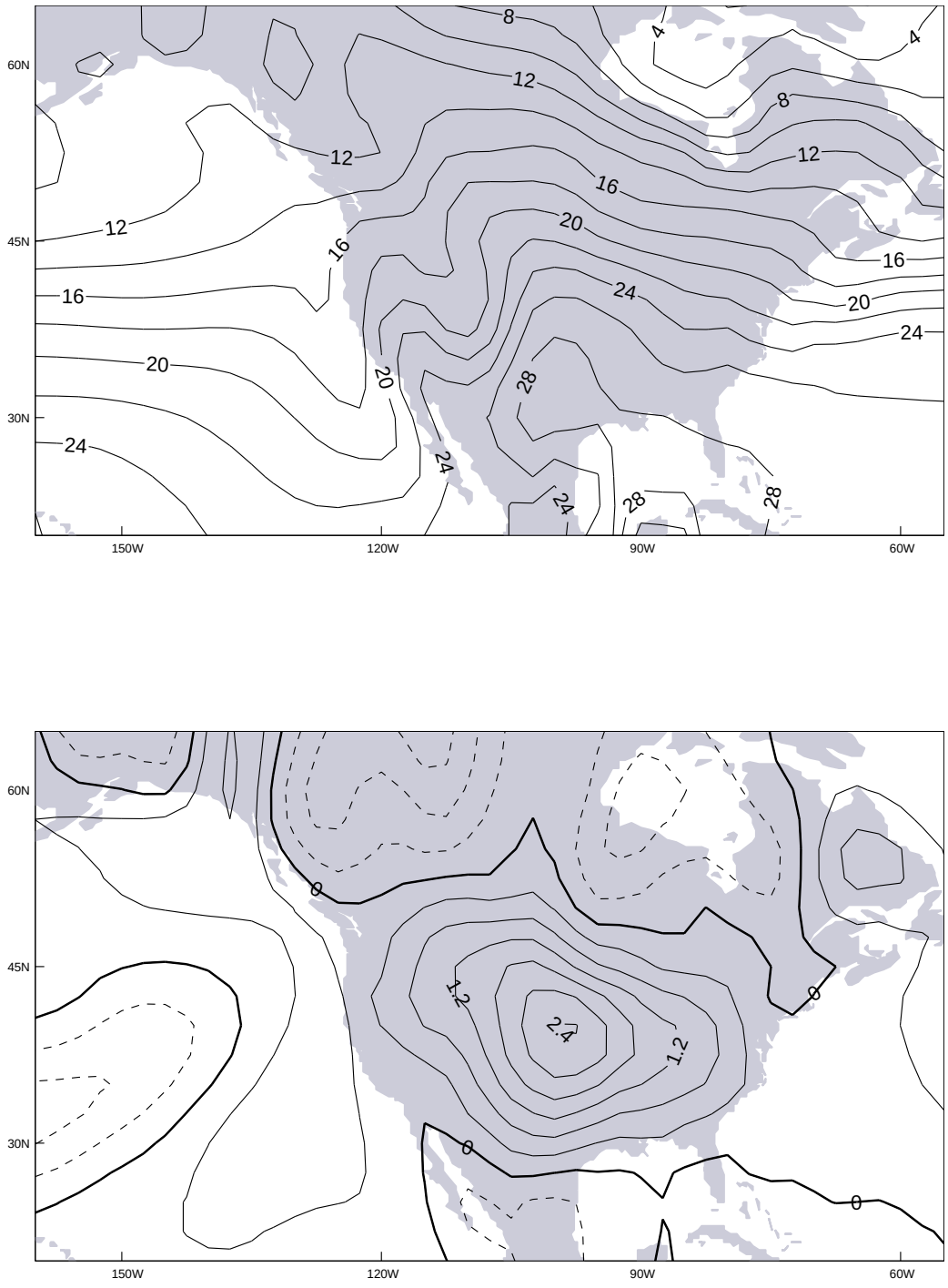

Figure 4.1: Surface air temperature averaged over three summers (JJA) for the years of 1934-1936 (top). Contour interval 2. Surface air temperature anomalies averaged over the same period (bottom). Contour interval 0.4. Units in $^{\circ}$C.

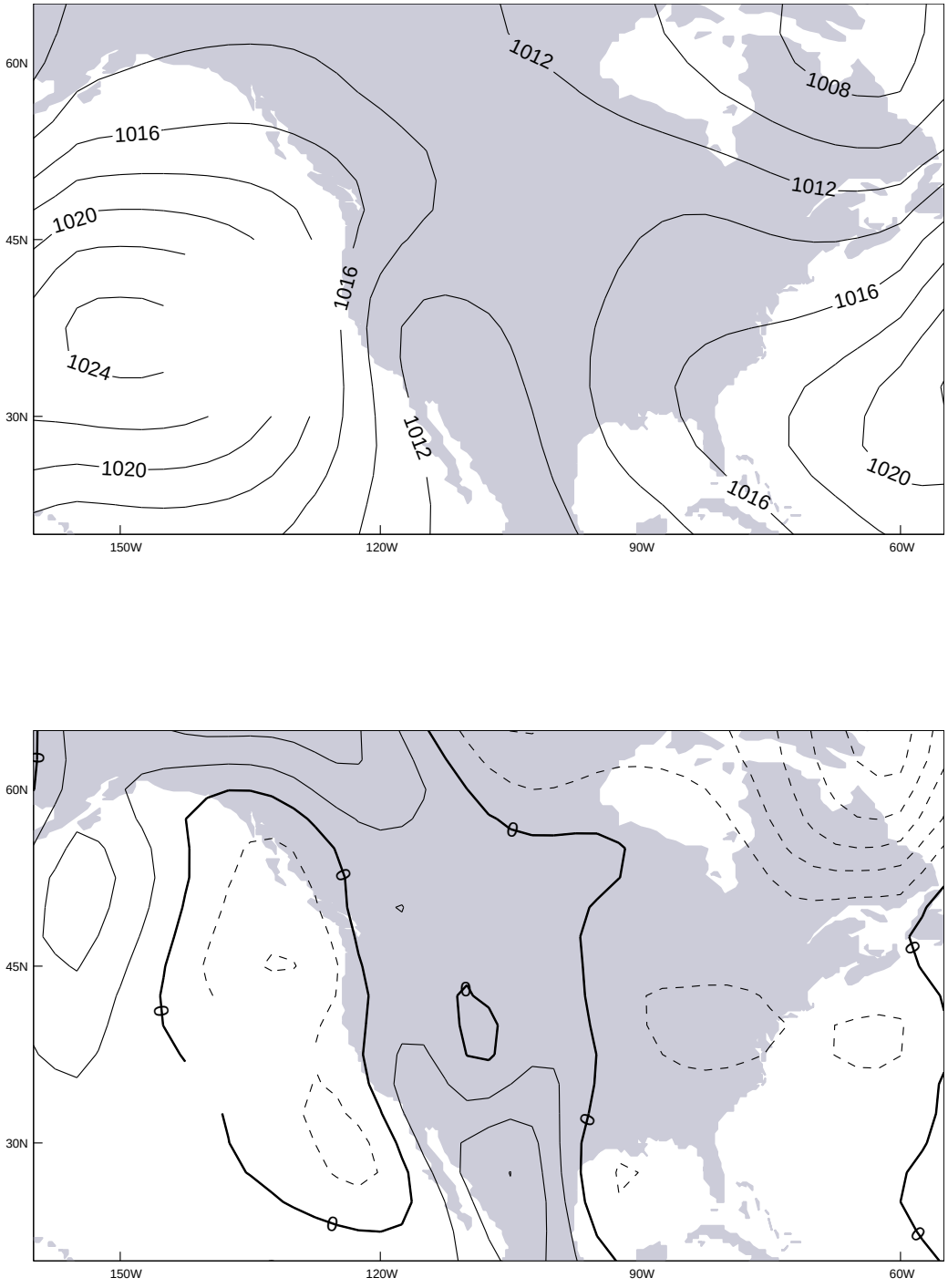

Figure 4.2: Sea-level pressure averaged over three summers (JJA) for the years of 1934-1936 (top). Contour interval 2. Sea-level pressure anomalies averaged over the same period (bottom). Contour interval 0.5. Units in mb.

Mexico up through central North America. Anomaly patterns show a region of above average heights over the Great Plains of the United States with the largest anomalies of $\sim$ 30 meters found over the area corresponding to the warmest temperature anomalies. The imprint of the SAT anomaly field is clearly discernible in the 500-hPa height anomaly field.

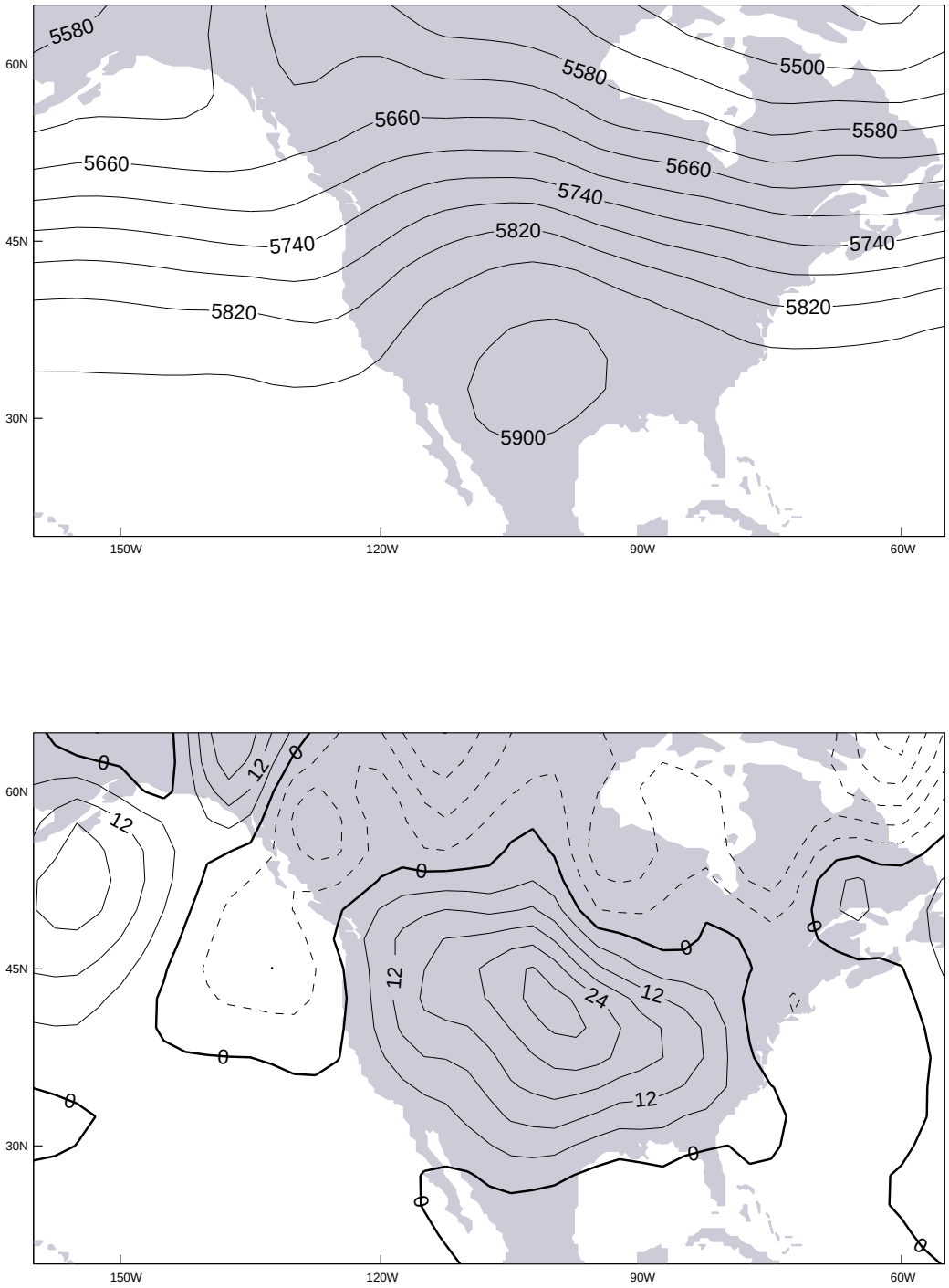

Figure 4.3: Reconstructed 500-hPa geopotential height field averaged over three summers (JJA) for the years 1934-1936. Contour interval 40. Reconstructed 500-hPa geopotential height anomalies for the same period. Contour interval 6. Units in meters.

Chapter 5

COMPARISON WITH OTHER STUDIES

5.1 Klein and Dai, Schmutz et al

The cross-validated results of this study, based on the independent data, from both the Western Hemisphere and Eastern Atlantic/Europe are compared with those of Klein and Dai (1998) and Schmutz et al. (2001) in this section. Klein and Dai's study was over the domain $20\,^\circ$ to $70\,^\circ$N and $180\,^\circ$ to $350\,^\circ$W, while Schmutz et al was on the domain $35\,^\circ$ to $75\,^\circ$N and $40\,^\circ$W to $45\,^\circ$E. Values are compared between the cross-validated datasets of this study to the cross-validated values of the other two (table 5.1). Results from those studies are for 700-hPa height reconstructions. Root mean squared error, computed using the unaggregated three winter and summer months of the reconstructed and analyzed datasets, and R^2 values of the cross-validated fields are used to compare the different schemes.

Table 5.1: Root mean squared error values for this study (500-hPa gph reconstruction) and Klein & Dai (1998) (700-hPa gph reconstruction) for western hemisphere poleward of $20\,^\circ$N. Comparison of results with Schmutz et al (2001) (700-hPa gph reconstruction) are for Eastern Atlantic and Europe.

Model Statistics					
		RMSE		R^2%	
Study	**Cross-validated**	**DJF**	**JJA**	**DJF**	**JJA**
Current	Yes	27	17	80	66
Schmutz et al (2001)	Yes	27	14	71	56
Current	Yes	26	15	69	52
Klein & Dai(1998)	Yes	30	24	87	74

Root mean squared errors from this study are slightly lower than those of Schmutz et al (2001) for the domain of Europe and Eastern Atlantic. Higher values of R^2 are realized

in this study. The root mean squared errors would be smaller were it not for the fact that their analysis is for the 700-hPa height field whereas ours is for the 500-hPa height field in which the variability is larger. The results of Klein and Dai appear to be inconsistent, both with Schmutz et al and with the present study. Root mean squared error for Klein and Dai are higher than in the other studies, yet R^2 values are also higher. One possibility is that the root mean squared error was calculated differently in their paper. Attempts to reconcile these inconsistencies were unsuccessful.

Chapter 6

CONCLUSIONS

6.1 Conclusions

It has been shown it is possible to reconstruct 500-hPa gph fields to a relatively high degree of accuracy using surface variables alone. The simple two predictor model, described in this paper, yields statistics comparable to the best published studies based on more complex procedures. This straightforward, pointwise regression scheme is highly robust, as evidenced by the very slight decline in R^2 when cross-validation is performed. A reconstruction of the 500-hPa height field during the mid-1930s using this model produces height anomalies consistent with weather conditions observed at the surface.

SAT and SLP explain a substantial amount of the 500-hPa height field variance. The leading EOFs and patterns of variance in the reconstructed fields agree quite well with those in the observations. Maps of regression coefficients for SAT and SLP derived from the pointwise scheme discussed here clearly reveals their effectiveness in reconstructing 500-hPa height as a function of latitude and longitude. SAT is generally more important over land while SLP dominates the model over the oceans.

The fidelity of the reconstructions, based on R^2, tends to be higher in winter than in summer and higher over polar and temperate latitudes than over low latitudes. Lowest correlation values are observed off the coasts of California and North Africa where SAT is decoupled from temperatures aloft. These regions tend to be characterized by low variability so they do not realize high root mean squared errors. The ability of the model to reconstruct 500-hPa height is also demonstrated through the high spatial correlations between the reconstructed and observed heights for the individual calendar months.

Reconstructions based on the NCAR/NCEP data are superior to those based on the independent data. This improvement is a result of the dynamical consistency between the upper level and surface fields in the reanalysis data produced by NCAR/NCEP. As a result there is already an amount of shared information between the different levels.

EOF analysis reveals the large scale patterns of variability of the analyzed 500-hPa height field are well represented by the model. The unexplained variance lacks coherent structure across the domain. In view of the relatively small scale of the error fields associated with the reconstructions, it may be difficult to improve on the results by applying corrections based on other statistical methods, e.g. maximum covariance analysis or canonical correlation analysis.

BIBLIOGRAPHY

Blackmon, M. L., Madden, R. A., Wallace J. M., Gutzler, D. S. (1979): Geographical Variations in the Vertical Structure of Geopotential Height Fluctuations *J. Atmos. Sci.*, 36:2450-2466

Jones, P. D., S. C. B. Raper, Bradley, R. S., Diaz H. F., Kelly P. M., Wigley T. M. L. (1986): Northern Hemisphere Surface Air Temperature Variations: 1851-1954. *J Appl Meteor*, 25(2):161-179

Kington, J. A. (1975): The Construction of 500-millibar charts for the eastern North Atlantic-European sector from 1781. *Meteorol Mag*, 104:336-340

Klein, W. H., Dai Y (1998): Reconstruction of Monthly Mean 700-mb Heights from Surface Data by Reverse Specification. *J Clim*, 11(8):2136-2146.

Namias J. (1955): Some Meteorological Aspects of Drought *Monthly Weather Review*, Sept 1955:199-205

Namias J. (1944): Construction of 10,000 Foot Pressure Charts Over Ocean Areas. *Bull. Amer. Meteor. Soc.*, 25:175-182

Schmutz, C, Gyalistros D, Luterbacher J, Wanner H (2001): Reconstruction of Monthly 700, 500 and 300 hPa Geopotential Height Fields in the European and Eastern North Atlantic Region for the Period 1901-1947. *Clim Res*, 18:181-193

Trenberth K., Paolino, D. A. (1980): The Northern Hemisphere Sea-Level Pressure Data Set: Trends, Errors and Discontinuities. *Mon Weather Rev*, 108(7):843-1086