The Development and Analysis of Climate Datasets for National Park Science and Management:

A Guide to Methods for Making Climate Records Useful and Tools to Explore Critical Questions

---

Report prepared for
National Park Service
Inventory and Monitoring Program

Final Report[1]
10 December 2009

Timothy Kittel[2]
Institute of Arctic and Alpine Research
University of Colorado, Boulder

[1] This document contains figures from copyrighted publications and is not for public distribution. Permissions pending.
[2] Contact information: kittel@colorado.edu / ph 303 258-0908 / INSTAAR, CB450, University of Colorado, Boulder, CO 80309

Table of Contents

1.0 Introduction

1.1 Objectives

Climate sets the stage for a landscape's species and ecosystems, albeit a moving stage. Temporal variability over a wide range of scales – from hourly events, year-to-year variation to decadal shifts and centennial trends – is a strong determinant of the state of populations and systems now and where they are headed. In addition, a site's climate variability tends to show strong correspondence with what is happening regionally and to dynamics at hemispheric scales. Consequently, understanding historical and current patterns in local climate and its links to regional to global processes is crucial for park monitoring, management, and research goals.

However, answering climate-related questions can be problematic. This can be because available climate datasets are often not ready 'off the shelf' for such applications without careful preprocessing or because appropriate tools for analyzing climate data are not well known or understood by users. Problems can also arise because nearby climate stations do not exist, or the necessary variables were not recorded for the period required. The goal of this report is to guide the development, analysis, and interpretation of climate data to address questions relating park resources and climate.

The objectives of this report are then:

(1) To layout a methodology for developing research-grade climate datasets appropriate for resource management science.

(2) To introduce the array of analysis techniques available for answering questions we often ask regarding climate dynamics and their interaction with landscape processes.

1.2 Approach

My approach is to identify common, key issues encountered when working with climate data and then suggest approaches and resources for pursuing specific solutions most likely to be useful to NPS personnel. The report is more guide than cookbook: for the specifics of any given process, I point to references to serve as an entry into the literature. In Figure 1, I lay out a workflow diagram that follows the overall components of this guide – the diagram portrays generic elements and decision points that are common to climate observation-based studies.

I start off with an outline for laying out a project's or program's data requirements as this will establish what problems are worth fretting over and what methods may or may not be appropriate (section §2.0, Figure 1a).[3] Second, I discuss methods for handling problems that typify station records (§3.0, Figure 1b). Next, I go through common analysis techniques and their interpretation (§4.0, Figure 1c). I wrap up with an overview of key considerations in the implementation of these processes (§5.0) – you may wish to review that synopsis early on to see the take-home messages.

For most situations, I draw on established statistical methods that are relatively common in environmental sciences and generally found in statistical software packages, and provide references where this is not so much the case. I cover caveats that go with these solutions and analyses. I give examples of some techniques in the figures, and make use of their captions to reveal detailed considerations needed to implement such methods.

[3] In the digital version of this report, cross-references to sections (§'s), tables, figures, and footnotes are hyperlinked. Websites linked to in the report were last accessed December 2009.

2.0 Establish Goals, Identify Requirements

Different research questions exploring climate processes have, of course, different data requirements (Figure 1a). Clearly establishing your research questions and then corresponding data requirements is key. Some requirements may be obvious, such as temporal and spatial specifications.[4] Others may be more subtle, such as what data problems need to be dealt with or can be ignored and which analytical techniques are appropriate. Creating a technically-fancy infilled daily dataset could be a lot of effort for little payoff if the goal is to look at centennial trends. On the other hand, such work can give greater confidence in evaluating trends in day-specific variables such as growing-season onset and termination.

2.1 Data requirements from hypotheses

Best practice starts with a clear *a priori* statement of hypotheses to be evaluated.[5] This is preferred over undertaking a large number of analyses in an attempt to find significant relationships. The downside of undirected numerous comparisons is there is a reasonable probably of one or more tests having statistically-significant results due to random chance alone.[6]

Your hypotheses will dictate the specific data needs of a research problem. Use their statement as a guide to identify:

- Variables of interest – what are the system drivers of consequence?
- Spatial scales of interest – single site vs. landscape or regional analysis, for example. (§2.2)
- Temporal scale – e.g., are daily event structure and the occurrence of extremes important, or just seasonal means? (§2.2)
- Record length and/or spatial density – to assure sufficient observations in time or space to reveal the patterns hypothesized.

2.2 Issues of time and space scale

Through this process, some forethought needs to be given to capture not only the multivariate but also multiscale nature of climatic controls on organisms and ecosystems. Selecting relevant temporal and spatial scales for analyzing meteorological data to reveal climate's impact can be guided by two principles:

Characteristic scales. Ecological processes tend to operate at characteristic temporal and spatial scales (Figure 2c, d; Delcourt et al. 1983, Urban et al. 1987). These initially prescribe climate analysis scales for given ecological processes. Note, however, there is not a 1:1 correspondence

---

[4] For example, spatial data specifications include spatial domain and station density or grid interval. Temporal specifications include record length requirements and timestep.

[5] See Schumm (1991: Chapter 2) for discussion of the scientific method in practice.

[6] Regarding such 'fishing expeditions' or 'data dredging' – If analyses use, for example, a probability threshold for Type I Errors[24] of $\alpha = 0.05$, then there is a theoretical likelihood that one out of 20 tests will give statistically-significant results due to random chance alone. If such multiple comparisons are part of a directed design, then this error can be controlled by setting an acceptable *experiment-wise error rate* and calculating a corresponding probability threshold for each comparison. This raises the bar for any comparison to be significant (i.e., lowers the probability threshold for each individual comparison). For more on multiple comparisons and experiment-wise error rate, see: Yandell (1997: §6.1) – available in part on Google Books (see References). Other multiple comparisons techniques are *familywise error rate* and *false discovery rate*. Another approach to avoid spurious results in any analysis is cross-validation, where analysis is re-tested on a random subset of data that were withheld from the original analysis.

between characteristic temporal/spatial scales for climatic and ecological processes (Figure 2b; see caption).

Scale interactions. While climatic and ecological processes have characteristic scales (Figure 2), factors controlling them operate across the range of scales. Climate forcing, for example, occurs through integration of finer effects and constraints of coarser ones.[7] A mismatch in temporal scales of aggregating station climate data (e.g., monthly values) can miss important relationships. Such is the case when the critical impact of weekly weather events on, for example, population processes is not captured by monthly data (e.g., Hallett et al. 2004; see also §4.4.2: *Process timescales differ, Mode 3*).

Spatially, a broader view of climate variability provides regional context to local climate forcing (§4.7) or, broader yet, an understanding of how hemispheric processes (e.g., El Niño) set the stage for local ecology year to year (§4.8). For the latter, Stenseth and Mysterud (2005) demonstrate the benefits of the span of spatiotemporal scales implicit in analyses linking hemispheric climate to local ecology (see §4.8.3: *Circulation indices broadly integrative*).

There are some common pitfalls in identifying suitable scales for aggregating climate data to reveal forcings on ecological systems. These include selecting the level of aggregation based on:

- What data are readily available – e.g., for temporal aggregation: monthly or annual means, when daily or hourly data may be more appropriate
- A poor understanding of ecological processes and their controls – either because a study is exploratory ('a fishing expedition') or established notions of how a system works are not well vetted

Such pitfalls are often difficult to avoid, but recognizing such limitations from the start can aid in hypothesis formulation, analysis choice, and interpretation. Timescale mismatching and aggregation issues are further discussed in the context of timeseries analysis, §4.4.2.

2.3 Data requirements of analytical methods

The next step is to identify what classes of analytical methods are needed to address your key questions and their data requirements. Are you interested in solely descriptive statistics (means, variances, frequency distributions; §4.1)? Do you as well aim to statistically test relationships either within the climate data or with other site variables (e.g., trend, correlation, and spatial coherence analyses; §4.3-4.8)? You may need to explore analysis options covered in these sections to identify corresponding requirements. Some general points on data requirements:

- Descriptive statistics require sufficient and unbiased sampling of the record to give representative results
- Statistical tests expect that certain assumptions about the data be met – such as having observations that are independent, normally (Gaussian) distributed, and identically distributed

In regards to the latter bullet, spatial and temporal climate data are characteristically not independent and some variables, such as precipitation, not normally distributed. I talk about these issues as they come up, yet keep in mind that some statistical methods are more forgiving than others in their requirements.

[7] Scale linkage through integration of finer-scaled effects and constraints of courser ones reflects the hierarchal nature of biological and geophysical systems (Allen and Starr 1982, Holling 1992).

In dataset development and data analysis sections (§3.0-4.0), I point out key requirements for many of the techniques I discuss.[8] Consult statistical references for specifics on these requirements, tests for these requirements, and possible workarounds (e.g., transformations). Some of this material can be found in on-line statistical texts and statistical software documentation (e.g., Helsel and Hirsch 2002, Schreuder et al. 2004, McDonald 2009, Garson 2009).[9] A highly-regarded reference for standard (parametric) tests (e.g., linear regression) is Sokal and Rohlf (1994) and for non-parametric tests (e.g., rank statistics) is Conover (1999), though there are many other useful texts along these lines.[10] Books focused on statistical methods in geophysical sciences include von Storch and Zwiers (2001),[11] Helsel and Hirsch (2002),[9] and Wilks (2006).[12] 'R' is a powerful statistical package which includes many of the techniques discussed – it is available free online.[13] On-line calculators for some tests include Kirkman (1996).[14]

2.4 Requirements guide what is possible with what is available

Together, data requirements dictated by your research hypotheses and by methods to test these will guide station selection and screening of the data (§3.0). For some research questions, available station data may not be up to the task. Alternatively, regional high-resolution gridded temporal climate data may do well for certain purposes (e.g., PRISM:[15] Daly et al. 2002, 2008, Di Luzio et al. 2008; Hijmans et al. 2005). Baron (2006) illustrates the use of a gridded dataset for point applications. Surrogate variables are also options for exploring climate processes of interest, such as surface hydrology to reflect watershed climates (e.g., Stohlgren et al. 1998) or station variable correlates (e.g., MTCLIM: Thornton et al. 2000, Running et al. 1987).[16]

---

[8] Of the techniques discussed in this report, linear regression is a one employed in many situations. I introduce it in the context of station change and missing data (§3.4.2-3.4.3, §3.6) and later in trend analysis and bivariate comparisons (§4.3.1, §4.4). If you use regression in other contexts, still refer to these sections (especially §3.4.3) regarding requirements, implementation, and pitfalls.

[9] General online resources include:
- Helsel and Hirsch (2002): http://pubs.usgs.gov/twri/twri4a3/html/pdf_new.html
- Schreuder et al. (2004): http://www.fs.fed.us/rm/pubs/rmrs_gtr126.html
- Lane (2007) – *HyperStat Online Statistics Textbook*: http://davidmlane.com/hyperstat/index.html
- McDonald (2009) – *Handbook of Biological Statistics*: http://udel.edu/~mcdonald/statintro.html
- Garson (2009) – *Statnotes:* http://www2.chass.ncsu.edu/garson/PA765/statnote.htm
- SPSS tutorial: http://www.stat.tamu.edu/spss.php

[10] Also:
- For linear regression, Draper and Smith (1998)
- Crawley (2002), a stats text with S-Plus applications
- Crawley (2007) is a reference manual for the R statistical language.[13] Also, see Verzani (2004) – available in part on Google Books (see References).

[11] von Storch and Zwiers (2001) available in part on Google Books (see References).

[12] Wilks (2006) available online on "Scribd:" http://www.scribd.com/doc/7128720/Statistical-Methods-in-the-Atmospheric-Sciences-Daniel-Wilks-

[13] R site: http://www.r-project.org/. A USGS course on R has useful materials online: http://www.fort.usgs.gov/brdscience/LearnR.htm, including resources listed at: http://www.fort.usgs.gov/brdscience/LearnR.htm#References.

[14] Kirkman (1996): http://www.physics.csbsju.edu/stats/; see also http://www.physics.csbsju.edu/stats/Index.html. Also: Wessa (2009) – http://www.wessa.net/rwasp_spectrum.wasp. These and other on-line tools are presented as possible resources, not as an endorsement or reflecting an assessment.

[15] PRISM is a 'smart interpolation system' that uses location-specific relationships among elevation, aspect, basin configuration, etc. and minimum and maximum temperature and precipitation to develop spatial interpolation functions.

[16] MTCLIM estimates solar radiation and relative humidity based on their physical relationships with minimum and maximum temperature and precipitation.

3.0 Methods for Making Climate Records Useful

3.1 Types of problems

3.1.1 Station issues

Once climate data requirements are laid out and station datasets identified as appropriate to these needs, the next task is to check for and deal with problems in these stations' records (Figure 1b). Nearly all station records have such problems, unless they have been extensively processed by another party (e.g., NOAA U.S. Historical Climate Network dataset, USHCN: Easterling et al. 1996, 1999, Endoe 2009)[17] – issues associated with processed datasets are discussed in the next section (§3.1.2).

Problems that typify station records are:

- Data errors – resulting in physically implausible values and questionable outliers (§3.3.1, §3.3.2)
- Collection biases (§3.3.3)
- Station changes – changes in location and instruments, changes in station environs (§3.4)
- Record length – whether sufficient for detecting temporal patterns (§3.5)
- Missing observations (§3.6)

These issues and common solutions are laid out in the sections indicated. I also discuss important tasks to do throughout, namely, to track data changes (§3.2) and document and evaluate the effects of your processing (§3.7). As you see from this list of issues to work through, creating a credible dataset suitable to your research questions requires an integrated, multi-stepped quality control (QC) process – some stages may be automated, but ultimately the process entails hands-on decisions (Peterson et al. 1998b: §5). Decisions are facilitated by involving those who are familiar with the region's climate and with the workings and limitations of QC tests.

As you evaluate the results of your efforts to create a usable dataset, keep in mind that individual station data problems are sometimes severe enough – too many missing values, intractable station changes – that estimated (corrected, infilled) values overwhelm original information, and it becomes more prudent to drop a station and rely on a nearby record.

3.1.2 Processed dataset issues

In the case of processed 'cleaned-up' data, your job is not over quite so easily. Understand what methods were used to handle these problems and evaluate if these methods are consistent with your needs. Judge if their techniques altered the data in a way that obscures a key process of interest, such as infilling missing days in mountain stations based an adjustment of valley records when understanding elevation contrasts are your research goal. Processing centers regularly flag modified datapoints and may offer versions with original data and data after different stages of processing so you can backtrack to the level of adjustment that matches your requirements (e.g., in USHCN Version 1).[18]

One class of processed climate data are area-averaged summaries. Area averages present their own issues. This is especially the case for climatically heterogeneous domains, such as U.S. Climate

---

[17] Easterling et al. (1996, 1999) constitute USHCN Version 1: http://www.ncdc.noaa.gov/oa/climate/research/ushcn/ushcn.html. Endoe (2009) provides a summary of the current research version (Version 2): http://www.ncdc.noaa.gov/oa/climate/research/ushcn/. See also §3.4.6 and footnote 59.

[18] Four levels are offered for USHCN v. 1 http://www.ncdc.noaa.gov/oa/climate/research/ushcn/ushcn.html#DATA

Divisions (Guttman and Quayle 1996, CLIMAS 2002).[19] Division dataseries are unweighted means of available stations at a given time. This means that available stations are not incorporated in proportion to how well they represent the domain's climate. Resulting problems include:

(1) As stations come into and leave the averages, the spatial representation within a domain shifts through the record. This is especially a concern over topographically-diverse areas.

(2) Averages of station records blend signals – the resulting signal's temporal variance is reduced and no longer realistically represents a region's climate. The degree of such signal blurring changes as stations come and go.

For these reasons, it is preferable to base analyses on individual station records. With these caveats in mind, however, areal averages can facilitate regional monitoring objectives (CLIMAS 2002, Wolter and Allured 2007). To reduce some issues with U.S. Climate Divisions, Wolter and Allured (2007) propose a revision with divisions based on climatic similarity.[20]

3.2 Version control and change flags

As you develop a workable dataset, document your process and provide means to undo your changes. Best practices dictate that:

- Adjusted, removed, or infilled observations be indicated with a flag
- Version control be exercised so that changes can be recovered, back to the original data if need be

Such tracking is crucial throughout the process of developing datasets and allows for evaluation of implemented changes. Include these elements in your documentation (§3.7).

3.3 Data errors, outliers, and biases

Data errors and biases come from observation collection, coding, and processing errors[21] and instrument and collection-protocol biases. These problems can be identified through screening for nonsense values (discussed next) and outliers (§3.3.2) and testing for known biases (§3.3.3).

3.3.1 Screening for reasonable values

Common quality checks for meteorological data are outlined in *NPS Climate Data and Monitoring Options* (Redmond et al. 2008); also refer to Peterson et al. (1998b: §2, App. A).[22] A good, generic guide to data checking is provided by Chatfield (1995: §6.4).[23] Screening can catch data transcription errors and common instrument errors such as dropouts, drift, biases, and other glitches if of sufficient magnitude (Figure 3). Such quality checks include nonsense and plausibility checks, such as a day's minimum temperature greater than the maximum and values beyond physical limits (see Table 1 for an example layout). Other checks look for questionable outliers and unusual behavior, such as spikes, step changes, or flat lines (§3.3.2).

---

[19] U.S. Climate Divisions are laid out more along watershed, economic, and political boundaries than by climatic (and so general ecological) similarity. Interactive maps of current conditions by U.S. Climate Division are at: http://gis.ncdc.noaa.gov/website/ims-cdo/div/viewer.htm

[20] Wolter and Allured (2007): http://wwa.colorado.edu/IWCS/archive/IWCS_2007_Jun.pdf, and see http://www.esrl.noaa.gov/psd/people/klaus.wolter/ClimateDivisions/.

[21] Be aware that some processing errors may include previous attempts to correct errors and you may need to 'undo' these (Peterson et al. 1998b).

[22] For a detailed review of errors and correction techniques for hourly data from automated networks, see Wade (1987).

[23] Chatfield (1995) – available in part on Google Books (see References).

Some data screening is done by data archive centers; otherwise, such processes need to be implemented as part of your own data protocol. You may decide that additional custom filtering of the data should be employed to meet specific needs of a project (cf. Redmond et al. 2008).

3.3.2 Outliers

Checks regarding unusual behavior can be among the most challenging to design. This is because errors can be hard to discriminate from real extremes and other meteorological dynamics. Any scheme to identify outliers as 'bad data' runs the risk of rejecting real events that have important ecological or hydrological consequences. Formally, we discuss this as a Type I error, that of rejecting good data.[24] Generally, climatologists are adverse to this error in favor risking a Type II error (accepting bad data) and so of retaining all outliers (provided they lie within physically-plausible limits). For your application, weigh the relative consequences of these errors to your analysis to decide how lax or aggressive to be in filtering out extreme values. NOAA regional climate centers, for example, differ in balancing these errors depending on their geography and prevailing user needs.[25] Outlier effects on regression analysis are discussed in §3.4.3.1.

Detection and rejection decisions for outliers or other unusual behavior are usually applied to daily or hourly data. These rely on a variety of methods that check for:

- Multivariate physical consistency – similar dynamics in physically-related parameters (e.g., among minimum and maximum temperatures, cloud cover, relative humidity, and precipitation)
- Climatological consistency – staying within the range of climatological norms, based on longterm record statistics
- Temporal consistency – rate of change tests looking for spikes and step changes
- Spatial consistency – similar dynamics in records of nearby stations

Detection processes can be automated – however, manual inspection is advised before flagged outliers are rejected. Meek and Hatfield (1994) present basic methods for (1) physical limits, (2) climatological filters, and (3) rate of change tests for daily and hourly records for an array of climate variables.

Multivariate physical consistency. For multivariate checks, Gandin (1988) gives examples of numerical tests and Redmond et al. (2008) of logic-based ones.

Climatological limits. Meek and Hatfield (1994) present a variety of climatological tests generally based on absolute ranges from the record (Figure 4c). This creates a rather strict test. More common climatological filters use some multiple of the standard deviation (SD) from the longterm mean to set limits. This multiple:

- Is usually large (e.g., >3 SD) – so more liberal than record limits

---

[24] More formally, a Type I error is the error of rejecting a null hypothesis (H_0) when it is *true* and should not be rejected. In this context, H_0 is that an outlier is from the same population as other observations (i.e., 'good data'). A Type II error is not rejecting a H_0 when it is *false* and should be rejected.

[25] For example, the Western Regional Climate Center (WRCC; Table 2) leans toward retaining all station data, including outliers which may or may not be true extremes. The High Plains Center (HPRCC), on the other hand, tends to eliminate likely outliers, with some possibility of tossing true extremes. These practices relate to common uses of the data in the respective regions – HPRCC data uses include crop models, which may be highly sensitive to daily extremes and so thrown off by outliers. WRCC's philosophy is that end-users can best determine criteria for handling outliers based on their specific application. Where these centers' regions overlap, data for the same station can differ depending on which center's dataset is accessed. (Stephen Gray, personal communications)

- Depends on climate variable – e.g., 5 SD for temperature (Figure 4d). For non-normally distributed data, such as precipitation, a multiple of the coefficient of variation (CV = SD/mean) is can be used; alternatively, data can first be normalized with an appropriate transform (§3.4.3.1) and then SD limits determined and applied in transformed space.
- May be asymmetrical – greater limits on one side of the mean than the other

Climatological limits also typically vary by season (Figure 4c, d) and can depend on site characteristics (for example, whether a site is maritime vs. continental, at high vs. low elevation, or in an air-shed prone to thermal inversion or cold-air drainage).

Peterson et al. (1998b: §4) review a variety of SD-based methods for detecting temperature outliers. SD thresholds may be staged – with a moderate limit (e.g., 2.5 SD) calling for additional evaluation, and a higher limit (e.g., 5 SD) outright rejection (Figure 4d). Box-and-whisker plots, described in §4.6.1, provide another frequency distribution-based method for identifying outliers. Longterm trends should be removed before implementing distribution-based techniques as trends can substantially contribute to a record's variance – it's more appropriate to judge outliers against a detrended record variance or vs. the local (shortterm) variance (Peterson et al. 1998b).

Temporal consistency. Some temporal inconsistencies, such as spikes and step changes, can be detected visually (Figure 4a) or numerically with rate of change checks (Meek and Hatfield 1994) (Figure 4b). The opposite case, too much temporal consistency – a flat line – suggests data dropout due to observer, instrument, or coding error.[26] Outliers and unusual behavior found in temporal tests can be further scrutinized with spatial consistency checks relying on neighbor comparisons (e.g., Eischeid et al. 1995; and Figure 4d caption). Used in tandem, these consistency tests bring both temporal and spatial perspectives to bear on judging outliers.[27]

Spatial consistency. In spatial checks, a primary consideration is how far away a neighboring station can be and still be relied on. This *spatial correlation length scale* depends on:

- Climate variable – spatial correlation generally drops off less rapidly for temperature than precipitation, especially for regions and seasons prone to convective rainfall (with a patchy distribution on hourly and daily bases)
- Time aggregation interval – spatial correlation generally extends farther for longer time averages (hourly vs. daily vs. monthly)
- Terrain uniformity – spatial length scale diminishes more quickly over heterogeneous terrain (e.g., regions that are mountainous, with significant landcover changes, or with varying degrees of lake/maritime influence)

Spatial tests are more readily applied in regions under a relatively uniform climate regime, such as the Great Plains, than one with sharp physiographic contrasts (e.g., the Rocky Mountains) or extremely sparse precipitation (e.g., the Southwest). These factors are discussed further in §3.6.5.

---

[26] Data flatlines can be identified as a series of observations with no or little change in value. These show up as an outlier in plots of short-period running standard deviations (like a running mean). Dropped observations should be marked as missing data. As an example test, implemented for PRISM input quality control for daily maximum and minimum temperature, a flatline period is identified as:

- ≥ 2 consecutive days with observations of 0.0ºC
- ≥ 10 consecutive days with observations within ± 0.4ºC.

5-10 consecutive days with observations within ± 0.4ºC are seen as a potential flatline period. These periods are statistically evaluated based on whether the 5-day moving standard deviation of the target station differs from that of surrounding stations; if differs and lower, then it is a flatline period. (Chris Daly, personal communication)

[27] A method that uses spatial comparisons to detect temporal inconsistencies is *double mass analysis*, discussed in §3.4.6 in the context of detecting station changes.

Hubbard et al. (2007) and You et al. (2007) evaluate spatial methods for detection of outliers in daily surface temperature and precipitation data, respectively; Eischeid et al. (1995) implemented a multiple-test spatial approach.

3.3.3 Collection biases

Data collection biases include those from:

Instrument calibration and design. Poor or drifting calibration issues can be accounted for if a close-by instrument can be used to develop a correction function or if a later re-calibration is well documented. Well understood biases associated with specific instrument changes are discussed in §3.4.5.

Time of observation for minimum and maximum temperature. For stations using minimum and maximum thermometers, time of observation other than midnight but rather at the more usual early morning or late afternoon times creates a bias in monthly mean temperatures. The bias can exceed 2ºC (Karl et al. 1986; Figure 5). This inconsistency arises largely from the time thermometers are reset relative to passage of warm and cold fronts.[28] This timing leads to cold biases in monthly means for morning reading stations and warm biases in means for late afternoon reading stations (Vose et al. 2003).

This is particularly troublesome for evaluating trends if observation time has changed during a station's history (see Figure 5 caption). A correction for this bias is presented by Karl et al. (1986) and further evaluated by Vose et al. (2003).

Missing observations that are time dependent (e.g., by weather, season, decade). Dataseries with dropped days that are pervasive, for example, at the start of a record or for periods of inclement weather will incur undersampling biases with time (problematic for trend analysis) or for bypassed synoptic conditions (problematic for daily event and seasonal analyses). Keep in mind the possibility that such temporal biases will interfere with questions being evaluated with these data. In general, these biases are not formally accounted for in studies – though they should be.

While missing days can potentially be infilled, the bias will persist in terms of lower variances: infilling estimation reduces variance in timeseries (see §3.4.3.3: *Reduction in variance*). Judging temporal biases with respect to infilling missing observations is discussed further in §3.6.2.

Multiday-accumulated totals for precipitation. Daily precipitation series suffer from the accumulation of precipitation over several days due to:

- Observer absence for manual precipitation gauges, resulting in several zero days followed by an accumulated value. These are typically flagged as an accumulated total.
- Sub-threshold input for tipping bucket gauges – If minor precipitation during a day doesn't accumulate enough to tip the collector but does with more inputs in the next day(s), then correct precipitation amounts will not be recorded on corresponding days.

[28] For stations with a morning observation time (e.g., 0700h), the passage of a warm front soon after thermometers are reset can result in the *minimum thermometer* retaining that 0700h temperature through the next 24 hours rather than capturing the next morning's pre-dawn temperature, the diurnal-cycle's actual minimum. For stations with late afternoon observations (e.g., 1700h), the passage of a cold front after thermometers are reset can result in the *maximum thermometer* holding that 1700h temperature through the next day, rather than catching that day's post-cold front max.

The only indication of an error will be if a nearby observer-read gauge shows distributed precipitation events during the same period.

When the data goal is monthly totals, multiday totals can be summed in. That is, unless the accumulation period crosses a month boundary. Then totals must be parsed. An accumulated total is often parsed out by day by apportioning the total guided by day-to-day amounts observed at a nearby station clear of these issues. This approach is likely to meet with success if the stations have good correspondence in daily structure (e.g., with tests in §4.6.1) and are likely to experience the same precipitation events (see §3.3.2: *Spatial consistency*, and §3.6.5: *Spatial heterogeneity,* regarding spatial correlation length scale). Such corrected data should be omitted in daily analyses (§4.6). If the total cannot be parsed in a defensible manner, then the total is dropped and all days included in the total are counted towards the months' missing day tolerances (§3.6.1).

An overview of manual station collection issues, such as for NWS cooperative network stations,[29] is presented by Leffler and Redmond (2004);[30] Daly et al. (2007) look at observer precipitation biases in detail.

3.4 Accounting for station changes

Station histories are rarely boring and tell of changes that give little consistency to observations. Troublesome station changes can as varied as:

- Station relocations to vastly different sites (even if nearby)
- New instruments with different response curves and accuracy (with no overlapping record)
- Dramatic adjacent landuse change – such as the introduction of irrigation, pavement, or an overshadowing building

Such changes introduce sharp, discontinuities into the record (Figure 6a). These changes are for the most part artificial, that is, 'non-climatic change points.' Temporal inhomogeneities arise as well from gradual changes, such as instrument drift (§3.3.3) or slowly changing conditions in the vicinity of stations (§3.4.7): for example, as landscaping or natural vegetation grows up and closer, and rural areas become urbanized. Peterson et al. (1998a) review many of these issues and numerical techniques to detect and handle them. Correction of temporal inconsistencies is important – ignoring them can have a major impact on climate change assessment (Figure 7). If, however, the intended use is for climate regime change detection, some detection and correction strategies are best avoided (see cautions in §4.5.3).

In this section, I discuss developing station histories as the first step toward accounting for station changes (§3.4.1) and then present several approaches for inhomogeneity correction depending on if:

- The station has overlapping records spanning the change point (§3.4.2)
- There is no overlap nor appropriate nearby station to guide adjustment (§3.4.4)
- The change is due to well-documented instrument issues (§3.4.5)
- There are reliable, nearby stations to guide adjustment – including cases where change points may be present but not sufficiently known from the station history (§3.4.6)
- Changes are gradual due to environmental changes (§3.4.7)

---

[29] National Weather Service (NWS) Cooperative Observer Program (COOP): http://www.weather.gov/om/coop/what-is-coop.html. Also referred as 'co-op' stations in this document.
[30] Online tutorial version: http://www.weather.gov/om/csd/pds/PCU6/IC6_2/tutorial1/Factors.htm

Section 3.4.3 gives a primer on regression – re implementation, interpretation, and cautions – as many of these approaches rely on this statistical method.

3.4.1 Building station histories

We have more confidence in numerical approaches for detecting and correcting temporal inhomogeneities if they are directed by knowledge of a station's history. Sussing out station changes can be an arduous process, however, and is one usually undertaken only if:

(1) Only a few stations are targeted.
(2) Their records are of high value with respect to a programmatic goal – such as for an important location.

– and –

(3) They bear a high probability of forming a high-quality record – that is, there's enough original information to form a strong foundation for developing a credible correction protocol.

To build a history for critical stations[31] –

- Consult station documentation to identify potential sources of temporal inhomogeneities, recognizing that not all changes get into such metadata.
  - For NWS co-op stations,[29] *Station History Reports* (B-44 forms) are available from NOAA regional climate centers and state climatologists (Table 2).[32] These reports describe station moves, instrument changes and major maintenance, and changes to the surrounding environment – though the reports are not necessarily complete.
- Review repeat photography, which can provide additional insights especially for gradually changing environs.
- Visit the site and its operator before undertaking a lengthy data protocol. This can provide an understanding not gleaned from station history reports and save effort in the long run. Intuition can be gained from:
  - Seeing station siting, instrument condition, and obvious recent environmental changes
  - Talking with operators, who very often have a longterm association with a station and can relate details that do not get into formal paperwork

If a location or instrument change is designed with an overlap between the old and new records, adjustments can be relatively straightforward, using, for example, linear regression (§3.4.2). Without overlapping records, a common method is mean offset and ratio adjustment (§3.4.4). While introduced here for simple cases, these approaches form the basic elements of more complex record processing for well-documented changes and other inhomogeneities (§3.4.5–3.4.7), merging sequential records of nearby stations (§3.5), and infilling missing observations (§3.6).

---

[31] For a guide to problematic station siting and observation practices to look out for, refer to Peterson et al. (1998), Leffler and Redmond (2004),[30] and Daly et al. (2007).

[32] Metadata describing station histories for NWS Cooperative and other U.S. stations are available from NCDC (http://www.ncdc.noaa.gov/oa/metadata/metadataresources.html) (see also State Climate Offices, Table 2).

3.4.2 Correcting known change points. I: The simple case with overlapping records – Regression

When there is sufficient overlap between outgoing (x) and new (y) station records, we can develop a linear regression conversion equation:

$$y'(t) = b_0 + b_1\, x(t) \quad , \tag{1}$$

where a value from the outgoing record $x(t)$, x at time (t), is adjusted to a value $y'(t)$ consistent with the new record.[33] The mark (′) indicates that the value is estimated; b_0 and b_1 are the regression y-intercept and slope, respectively, determined from the observed (x,y). An example correction based on overlapping instrument records is shown in Figure 7 for a precipitation gauge change – the correction had a substantial impact on a longterm trend analysis.

Details in implementing regression are covered next (§3.4.3).

3.4.3 Regression analysis primer – Implementation, interpretation, caveats

While I'm presenting this primer on regression embedded in the section on correcting station changes, these guidelines apply to using regression with climate series in general. I'll refer back to this section for other applications later in the report.[34]

3.4.3.1 Regression analysis implementation details –

I discuss in this section considerations in implementing regression analyses – these are with respect to:

(1) Timescale considerations
(2) Numerical method
(3) y-intercept method

and the means to deal with:

(4) Autocorrelation
(5) Outliers
(6) Data distribution requirements
(7) Nonlinearity

Time-related considerations: Record overlap, timestep, and season dependence. There will be more confidence in developed conversion equations:

- For cases with longer overlap periods
- If the conversion timestep is at the original collection temporal resolution – e.g., daily rather than on temporally aggregated values. Conversions at daily or weekly rather than monthly or longer time intervals are better able to capture nonlinear relationships. On the other hand, there is little expectation that hourly data are correlated between sites except at the closest distances.[35]
- If conversion relationships are season or weather-regime dependent. Conversions that account for seasonal shifts in overall weather regime better reflect relationships between

---

[33] It doesn't matter which of the early or current record is adjusted – in the example implementations here and §3.4.4, the earlier series is adjusted to match the current record.
[34] Other sections that deal with regression in some detail are §3.6 (*Missing observations*), §4.3 (*Trend analysis*), and §4.4 (*Covariation among variables*)
[35] Recall that station moves are not necessarily over short distances (§3.4).

locations. One technique is to base analyses on a moving window (in time, e.g., spanning 28 days) to avoid sharp temporal boundaries in conversions.[36]

Numerical implementation. Overlapping observed records (*x, y*) are used to calculate b_0 and b_1 in eq (1) using best-fit techniques:

- *Ordinary least squares* (OLS) is the most common method for determining the equation that best fits the data – and is the primary technique provided in most statistical packages and presented in statistics books (cf. §2.3).[37] OLS assumes that observations are *independent, normally distributed,* and *variance constant* (homoscedastic).[38]
- *Robust regression* methods, while less powerful than ordinary least squares, are alternate techniques that are 'robust' with respect to the limitations of non-normal or heteroscedastic data, as well as 'resistant' to the effect of outliers (discussed later in this section).[39]

Ordinary least squares regression is the method of choice (in terms of testing power) when data distribution assumptions are reasonably met directly or with data transformations (discussed shortly), while robust regression techniques are more general and can be applied with fewer restrictions (Wilks 2006). In the heteroscedastic (but still normal) case, an appropriate robust regression method is *weighted least squares* (WLS; Helsel and Hirsch 2002: §10.3).[40] In WLS, observations (*y*) are weighted by the inverse-square root of their local variance[41] – this accounts for variance changes along the range of *y*.

Zero intercept. Depending on the nature of the relationship between two records (*x*,*y*), you may force the *y*-intercept b_0 to zero in setting up the regression analysis. While this may make sense logically, in practice it generally does not produce the best results. This is because, when not set to zero, b_0 is the regression line's *y*-offset determined across the full range of data and so is not so much about what's happening at $x = 0$.([42])

Serial correlation. Climate data are typically serially correlated – that is, not independent. I discuss tests and corrections for serial correlation in §4.3.1 in the context of trend regression analysis – these are applicable here as well.

Outliers and end members. Outliers can have an undue influence on a regression analysis, especially for ordinary least squares regression. Outliers near the end of a regression line (end members) have significant leverage in determining the slope (b_1) of the line.[43] To check for these issues:

---

[36] In this approach, the conversion for any given day of the year is based on all daily data in the record that fall within, for example, a 28-day window centered on the day.

[37] von Storch and Zwiers (2001: §8.3.15-18)[11] review pros and limitations of the least squares method.

[38] See Helsel and Hirsch (2002: §4.4)[9] re tests to evaluate whether data are normal. See also Steinskog et al. (2007). Residual plotting techniques for evaluating heteroscedasticity are discussed in §3.4.3.2.

[39] Re: robust methods –

- Wilks (2006: §3.1.1)[12] presents a general discussion on robustness and resistance.
- Helsel and Hirsch (2002)[9] cover robust and resistant methods throughout their book; robust regression techniques are treated in Chapter 10.
- Software packages – robust techniques are offered in R and SPSS, for example.

[40] Helsel and Hirsch (2002: §10.3)[9] present a technique for performing WLS using OLS linear regression software.

[41] *Local* variance refers to variance of *y* determined for an interval along the range of *y*.

[42] If a regression line substantially misses the mark for observations near $x = 0$, a nonlinear transformation may be in order (covered later in this section: *Transformations for modeling nonlinear relationships*).

[43] See Helsel and Hirsch (2002: §9.5.1)[9] re: outliers.

(1) Graph y vs. x (scatter plot) to reveal suspect points
(2) Check that outliers are not errors (§3.3.2)
(3) Test the sensitivity of the regression to outliers by re-running the analysis omitting them
(4) If the regression is sensitive, apply a robust technique[39]

Transformations – to adjust data distribution. Transformations may be used to adjust data to normality (or at least to symmetry) and constant variance.[44] Precipitation data, for example, are rarely normally distributed, as are variables reported in relative proportion or percentage units (relative humidity, % sunshine). Common precipitation transformations include logarithmic, square root, and cubic root.[45] A transformation of proportional data is arcsin[square root(x)], where $0.0 \leq x \leq 1.0$ (McDonald 2009).[46] Wilks (2006: §3.4.1)[12] lays out a process for determining the most appropriate power transformation for a given dataset.[47] For issues arising from transformations in regression analysis, see §3.4.3.3.

Transformations – for modeling nonlinear relationships. If you have reason to believe that the x, y relationship is nonlinear, examine the data initially with a y vs. x scatter plot. Based on the shape of the plot, explore straightforward nonlinear transformations of either or both x and y.[44,48] Re-graph to see if the conversion had the desired effect. The resulting 'linearized' (x, y) is then used in the linear equation (eq 1). As just noted, for concerns re transformations in regression, see §3.4.3.3.

3.4.3.2 Judging results –

To decide if the regression equation you've developed is sufficient to the task:

(1) Check the regression significance level (p value)[49] – Reject the result if p is not within an acceptable level (e.g., $p < 0.01$, or at least < 0.05).

(2) Judge the predictive power of the regression – The regression coefficient (= coefficient of determination), R^2, can be interpreted as the proportion of the dependent variable's (y's) variance that is explained by the regression equation. Ask if it is high enough to be useful (e.g., $R^2 > 0.60$ = 60% variance explained). *N.B.,* it is not sufficient to have high significance (a high p), as this does not necessarily mean that a large amount of the observations' variance is explained.[50]

(3) Visually check a scatter plot of the regression's residuals (y'-y) vs. x – where y' and y are the predicted and observed value of the dependent variable in eq (1). A residual plot can reveal problems in the data or regression model used. If the residual plot shows a pattern other

---

[44] Resources re transformations:
- Helsel and Hirsch (2002: §1.7, §9.3, §9.6)[9]
- Wilks (2006: §3.4.1) for more specifics[12]
- von Storch and Zwiers (2001: §8.6.2)[11] present complex functions to resolve common issues.
- McDonald (2009: pp. 160-164) = http://udel.edu/~mcdonald/stattransform.html.

[45] Examples – (a) to apply a square-root transformation in eq (1), both x(t) and y'(t) would first be converted to SQRT (precipitation). (b) see eq (5) in §3.4.4.

[46] For percentage data, first convert x to a proportion: x/100.

[47] Power transformations, as some of those just noted for precipitation, include power of x (e.g., square-root $x^{1/2}$, x^2, inverse x^{-1}), exponential (e.g., e^x, e^{-2x}), and logarithmic [$ln(x)$].

[48] Many statistical and spreadsheet packages have nonlinear line fitting routines that can aid in exploring possible x-transformations, such 'trendlines' in (Microsoft Office) Excel (for trendlines options, select: display equation, display R-squared).

[49] The p-value is the probability that the relationship found is due to chance alone, rather than to a hypothesized relationship. More formally, this is the probability of committing a Type I Error (see footnote 24). von Storch and Zwiers (2001: §4.1.7, 4.1.9-.11)[11] review the interpretation of statistical tests.

[50] The interpretation of p vs. R^2 is also discussed in the context of teleconnections (§4.8.3: *Prediction*).

than a random normal, homoscedastic dispersion from a zero-slope line, then it is likely that transforms to adjust for data distribution issues or nonlinearities are needed or that those already applied need to be reassessed.[51] Wilks (2006: §6.2.6)[12] lays out an approach for diagnosing residual plots.

If the significance test's p, R^2, or residual plot are not satisfactory, review implementation details (§3.4.3.1) and pitfalls (discussed next, §3.4.3.3); see also Helsel and Hirsch (2002: §9.5)[9] re regression diagnostics. Other caveats in interpreting regression results are given in §4.4.2 (for the case of regressing timeseries of two different variables)[52] and §4.9 for statistical results in general.

3.4.3.3 Pitfalls in linear regression results –

Before relying on regression results, take into consideration two possible analysis artifacts: (1) variance reduction arising from statistical estimation and (2) inconsistencies arising from transformations.

Reduction in variance. A problem stemming from using regression-estimated values in a dataset is that these estimates reduce the variance of a timeseries. This is because the conversion equation delivers the best estimate, but without accounting for randomness associated with observation. Such reduced variance will cause standard statistical tests you implement in later analyses to overstate the significance of your results. If a fair portion of the data is estimated (from this and other preprocessing steps), keep in mind the consequences of this artifact when interpreting your final results. More sophisticated regression methods include an 'error' term in equation (1), which stochastically adds back in such noise to predicted values.

Transformations and R^2. In regression, transforming the *dependent* variable y – whether to adjust data distributions or to linearize the model – complicates the interpretation of R^2. This is because the transformation alters the variance structure of the underlying data, such that R^2 becomes the variance explained relative to the *variance of the transformed y*. As a result, R^2's for regressions with linear and different transformations of y are not on comparable scales (Scott and Wild 1991) – that is, you cannot judge whether a transformation produced a better regression equation based on R^2's.[53]

Kvålseth (1985) and Scott and Wild (1991) suggest that regressions with transformed y are best evaluated on how well predicted y' match observed y back in their original scale, rather than by the regression R^2 in transformed space. This is done by back-transforming predicted y' and observed y, and regressing (y', y) pairs now in linear space to determine the appropriate R^2 – one that relates to the original data and is comparable across models. The resulting overall strategy is then: the best-fit *regression equation* (and its statistical significance, p) is determined in y-transform space, but the R^2 used to evaluate % variance explained is calculated in the original (linear) space. Willett and Singer (1988) discuss this issue with respect to weighted least squares robust regression (WLS, §3.4.3.1).

For other scale disconnect issues regarding back transformation of data distribution parameters (e.g., mean, SD), see §4.6.1.

---

[51] Helsel and Hirsch (2002: §2.3.3)[9] note difficulties in visually judging residual plots for heteroscedasticity, and present a smoothing method to avoid these problems.
[52] under: *Interpretation of correlation and regression results* in §4.4.2
[53] See also Helsel and Hirsch (2002: §9.6).[9] This is an issue only when the dependent variable y is transformed, not if just independent x's are transformed.

3.4.4 Known change points. II: The simple, single station case, without overlapping records – Offset and amplitude adjustment

The regression method is obviously not possible if there are no overlapping records associated with a station move or instrument change. If, on the other hand, there is a relatively clean step change in a climate variable with the station change (e.g., Figure 8), then an easy method is to adjust one part of the record or the other by such a step offset.[33]

Mean offset adjustment. For the case of adjusting the early record (x) to be consistent with the more recent record (y), the offset is simply based on the difference in the record's *shortterm means* on either side of the change point, $\Delta_{y\text{-}x} = (\bar{y} - \bar{x})$. Each observation in the early record, x(t), is then shifted by the offset:

$$x'(t) = x(t) + \Delta_{y\text{-}x} \quad , \tag{2}$$

where x' (t) is the adjusted value of the early record consistent with the more recent record y. The means ($\bar{y}$, $\bar{x}$) should be for a period long enough to represent the two parts of the record, but not so long that their difference is affected by decadal shifts in climate regime or longterm trends.

Variance adjustment. If the variance of the record also changes at the change point, then the early record (to continue the case just used) can be *amplitude* adjusted by the ratio of the standard deviations of the two records. Amplitude adjustment, however, is applied to *deviations* of the early record from its mean, $\delta\, x(t) = x(t) - \bar{x}$. This creates an anomaly timeseries.

The amplitude of the anomalies is then adjusted by the ratio of standard deviations after vs. before the change point:

$$\delta\, x'(t) = \delta\, x(t) \times (SD_y / SD_x) \tag{3a}$$

$$x'(t) = \delta\, x'(t) + \bar{y} \tag{b}$$

In eq (3a), $\delta\, x'$ (t) is the adjusted anomaly series and SD_x and SD_y are standard deviations over the same shortterm periods used to calculate $\bar{x}$ and $\bar{y}$. The corrected timeseries is constructed from anomalies by adding back in the mean: for the adjusted early record (x') to be fully consistent with the recent period, we add in the mean for the recent period, $\bar{y}$ (eq 3b). The process of taking out $\bar{x}$ at the start and then adding $\bar{y}$ at the end incorporates the offset adjustment in eq (2).

Considerations for specific variables –

Minimum and maximum temperature. Offset adjustment is often used for temperature series. Where both minimum and maximum temperatures (T_{min}, T_{max}) are being adjusted, the physical requirement that $T_{min} < T_{max}$ can end up being violated when eq (3) is independently applied to these linked variables. A common approach for avoiding this issue is to recombine them into the generally independent variables mean temperature (T_{mean}) and diurnal temperature range ($DTR = T_{max} - T_{min}$). We then apply record adjustment techniques to the derived variables, and subsequently restore the adjusted timeseries to T_{min} and T_{max}.[54]

Precipitation. Because the frequency distribution of precipitation is highly skewed, a nonlinear adjustment is more appropriate. A simple technique is scaling the earlier record x(t) by the ratio of the means:

[54] Adjusted records of T_{min} and T_{max} are reconstructed from T_{mean} and DTR as: $T_{min} = T_{mean} - ½\ DTR$, and $T_{max} = T_{mean} + ½\ DTR$. Note that, alternatively, DTR and T_{mean} can be analyzed in place of T_{min} and T_{max} (§4.2.1).

$$x'(t) = x(t) \times (\bar{y}/\bar{x}) \quad (4)$$

In a more generalized approach, we use a nonlinear transform to try to make the precipitation frequency distribution function roughly symmetric about the median. Select the transform – such as a log, square root, or cubic root transform – that best suits the data (§3.4.3.1: *Transformations*). Using a natural-log transformation, eq (2) would be:[55]

$$ln[x'(t)] = ln[x(t)] + [ln(\bar{y}) - ln(\bar{x})] \quad (5)$$

3.4.5 Known change points. III: Well-documented instrument changes

Some instrument changes are so common they have been well studied and correction practices recommended in the literature. Precipitation gauges vary in catch dynamics and their biases, such as for wind-driven undercatch of snow, are well documented. Corrections methods are described by Legates (1995) and Yang et al. (2005).[56] For NWS co-op station temperatures, a change over in the mid-1980's from liquid-in-glass thermometers to the electronic sensor Maximum/Minimum Temperature System (MMTS) introduced a bias in these records (Figure 8). The bias and its adjustment are examined by Quayle et al. (1991; see also Doesken 2005). Biases due to change in time of observation at manual sites are also well documented – this bias and its correction are discussed in §3.3.3.

3.4.6 Seeking and correcting undocumented inhomogeneities – Role of neighboring stations

Most station changes, however, are not accompanied by overlapping records nor are well documented. In this case, the common method relies on examination of records of nearby stations to detect and correct timeseries discontinuities in the station of interest (Karl and Williams 1987, Peterson et al. 1998a).[57] Such techniques can also be used to guide and confirm correction of known heterogeneities (§3.4.2–3.4.5).

The general procedure is:

(1) Locate potential record shifts using station histories as well as possible (§3.4.1)
(2) Detect and evaluate potential inhomogeneities by comparison to a climate reference timeseries
(3) Adjust the station record if needed, guided by the reference series

A *reference series* can be based on a single neighboring station or a network of stations strongly correlated to the station being evaluated. For such corrections to be successful, a key criterion for a reference series is that it is 'clean' (temporally homogeneous) so that changes in its record only reflect the climate. As this is rarely the case, Peterson and Easterling (1994) give a method for optimizing the creation of reference series. Because metadata are often incomplete, Menne and Williams (2005) present more sophisticated methods that identify inhomogeneities in records without station history guidance. Enloe (2009) presents an overview of these protocols as used in USHCN Version 2.[17]

---

[55] Note that eq (5) can be re-expressed as: $ln\ [\ x'(t) = x(t) \cdot (\bar{y}/\bar{x})\ ]$, which reveals the relationship in eq (4).
[56] Photographs showing the variety of precipitation gauge types: http://www.uaf.edu/water/faculty/yang/bcp/photos.htm. Gauge biases are illustrated in online images from Yang et al.: e.g., Northern Hemisphere January measured and corrected precipitation: http://www.uaf.edu/water/faculty/yang/bcp/pm_jan.gif and http://www.uaf.edu/water/faculty/yang/bcp/pc_jan.gif, respectively.
[57] Another approach uses breakpoint analysis (Haimberger 2007, Christy et al. 2009). Another application of breakpoint analysis is discussed later re regime shifts (§4.5.3).

Temporal inconsistencies at a station relative to neighboring stations can also be identified and corrected with *double mass analysis*. This technique plots accumulated station values over time against the accumulation for a reliable neighboring station or the average accumulation across a set of stations. The line should be straight with little variation if records are consistent; any breaks in slope point to a systematic change in a station's collection regime (Kohler 1949).[58]

3.4.7 Station environment changes

Once non-climatic change points have been accounted for in a dataseries, some local human-induced gradual changes in climate may still remain, such as from changes in environs. For example, an urban heat island effect is found in timeseries for stations once in rural environs and now urbanized. Accounting for such warming effects is explored by Karl et al. (1988; Figure 9).[59] The climatic effects of other land use change can also be imbedded in the records of stations near land converted from, for example, forest to pasture and then to irrigated crops. Such changes can locally affect mean temperature, diurnal temperature range, atmospheric moisture, and precipitation – Hale et al. (2006) explores some of these impacts in station records.

If these changes might hamper detecting the signal you're interested in, it is important to remove or at least document these effects. However, keep in mind that these gradual climate changes are real even if they only reflect highly localized effects and not regional dynamics. Removal is warranted only if they interfere with addressing your research questions.

3.5 Concatenating timeseries of nearby stations to extend the record

To extend a site's climate history farther back in time than its current record, we can concatenate (link) its record with that of an older, similarly behaving station in the region (Figure 6). Similarly, if a station of great value (e.g., because of location and longevity) has been dropped from the observation network, we can substitute in an ongoing record from a nearby station. Concatenation can be completed much in the same way as we handle known station discontinuities (§3.4.2-3.4.4).

Generally, first evaluate quality and comparability of the two records. This entails:

- Ensuring that both records are temporal homogeneous (§3.4)
- Testing for a strong correlation between stations over a reasonable period of overlap – If the overlap is sufficiently long, similarity in station behavior at multiyear and longer timescales can be evaluated using spectral analysis (see Figure 6 caption and §4.5.2).

Second, concatenate the two series by:

- Adjusting one of the two series to conform to the other (§3.4.2-3.4.4)
- Choosing a point to switch from series to the other – This is often broadly based on judgment as to which original series is of higher quality and most important to preserve and fine tuned so as not to create a local discontinuity at the switch-over point.

---

[58] NWS/NOAA user's guide for a Unix program for double mass analysis provides an overview of the technique: http://www.nws.noaa.gov/oh/hrl/idma/html/dma_home_frame.htm

[59] Urban heat island effects are by their nature automatically treated in the scheme of Menne and Williams (2005), referred to in the previous section (§3.4.6); see Endoe (2009).[17]

3.6 Missing observations

When data are missing, three options are:

(1) Decide its ok to ignore in aggregations (§3.6.1)
(2) In the analysis phase, choose a statistical technique that allows for missing values (such as the Mann-Kendall trend test, §4.3.2)
(3) Infill them (§3.6.2)

3.6.1 Aggregation and missing day tolerances

If you have daily data and will be analyzing monthly values, missing a few days per month can be acceptable. An approach is to allow some number of days of missing precipitation data (e.g., 3 days/month) and of missing temperatures (e.g., 5 days/month) for the month to be considered complete. Generally, missing temperature values are of less consequence than missing precipitation because a few days of precipitation can account for most of a month's accumulation. For some climate regimes, such missing-tolerance levels may need to be far lower (e.g., sites prone to rapid Arctic frontal passage, summer rain primarily as convective storms) than for others (e.g., maritime climates with low day-to-day variability).

3.6.2 Infilling – Initial considerations

Infilling missing data should be processed last, after data errors, biases, and inhomogeneities are taken care of.[60] First, evaluate how much of the record is missing – and if these gaps are random across the record or temporally biased (discussed earlier, §3.3.3: *Missing observations that are time dependent*). It is often sufficient to judge such biasing visually (e.g., Figure 10).

The importance of missing values and temporal biasing depends on analysis requirements. If you plan to use a statistical technique that permits missing data, then infilling may not be needed provided that gaps are not huge and their temporal distribution not particularly biased (e.g., Hirsch and Slack 1984). In deciding whether to infill, also consider previously mentioned cautions re the effect of estimated data on variance structure (§3.3.3, §3.4.3.3: *Reduction in variance*).

Also keep in mind you may need to treat missing values differently for different analysis goals. Infilled dailies may be required for creating a reliable monthly dataset – e.g., for calculating trends or as input to a resource model. On the other hand, infilled values will interfere with daily frequency distribution and extreme value analyses (§4.6). In this case, conflicting requirements can be handled using the same dataset for both purposes by flagging infilled data for omission when called for (§3.2); in other cases, a few versions of the dataset may be called for to meet diverse analysis goals.

In the next sections, I present several different infilling approaches – including statistical and non-statistical spatial models (§3.6.3–3.6.6) and temporal statistical models (§3.6.7).

3.6.3 Infilling using a single or few neighboring stations – Simple regression models

Spatial methods (covered here and §3.6.4) are powerful techniques for infilling missing values. Much like spatial techniques for correcting inhomogeneities, these methods evaluate and utilize correlations between a target station (with missing data) and its neighbors. The most

[60] Concatenation (§3.5) can be done either before or after infilling. Concatenating before gives the benefit of having a longer record on which to base infilling, but the disadvantage that infilling relationships are based on a derived record for part of the period.

straightforward technique uses standard linear regression analysis. In a simple implementation, the steps are:

(1) Select candidate nearby stations based on:

- Proximity
- Similar environmental controls over climate
- Completeness of record at times when values are missing at the target station
- Sufficient record length on which to build a statistically significant regression

(2) Run separate, simple regressions[61] (eq 1; §3.4.2) of the target record on each candidate station record covering a specified period. The specified period can be defined using various criteria – the objective is to optimize the predictive power of the regression model for a given situation. Examples are:

- Use the entire record (all years) for the corresponding season or month (e.g., all daily data across all Januaries)
- Use a moving window about a missing value's date (such as daily data from 2 weeks on both sides of the date for that year) – An advantage of this approach is that the regression is tuned to weather conditions happening around the missing date. Note that which neighboring station is selected as the best predictor can change as the window is moved to each missing point.

(3) From these regressions, select the single station with the greatest predictive power (highest statistically significant R^2). Follow the guidelines for accepting a regression model laid out in §3.4.3.2.

Refer to caveats and considerations regarding regression-based estimates in §3.4.3.3 and re spatial models in §3.6.5. If this method does not produce a workable model, explore using:

- A multivariate regression model – Following the single-station procedure, but with terms for additional stations in eq (1).
  - Note: Add in a limited number of other stations only to the extent that R^2 is significantly improved. Models with too many predictor terms, each adding small improvements in R^2, are prone to being 'overfit' (where noise is being modeled as if it was true signal) giving false confidence in the model's predictive power.
- Spatial autocorrelation models – §3.6.4
- Alternative, non-statistical spatial schemes – §3.6.6
- Temporal models – §3.6.7

3.6.4 Spatial autocorrelation models

More sophisticated infilling techniques involve modeling a correlation surface through space and using this surface for prediction at a point – the target station. The spatial autocorrelation surface is created based on a relatively large network of stations (e.g., 10-100 stations) that are locally well correlated with the target station. Geostatistical techniques include kriging (Haas 1995) and thin-plate spline prediction models (Hutchinson 1995, 2004); the development and evaluation of a kriging model is illustrated in Figure 11 and Figure 12 (discussed in the next section, §3.6.5). A classic reference for spatial statistics is Cressie (1993a).[62]

---

[61] Simple regression is with only one independent variable, x (in this case, a neighboring station), in comparison to multiple regression with two or more independent variables.

[62] Another text is Clark and Harper (2000); the 1979 edition (by Clark) is online at: http://www.kriging.com/PG1979/ (as pdf: http://www.kriging.com/PG1979/PG1979_pdf.html).

In applying these methods, it is usually more effective to work with anomaly (means removed) or standardized anomaly fields (anomaly divided by standard deviations) than with original values.[63] This is because temperature and precipitation anomalies tend to be regional in scope and so more spatially coherent than original fields. Additional techniques simultaneously evaluate spatial and temporal autocorrelation to infill missing data (e.g., Kondrashov and Ghil 2006, 2007).[64] Interpretation of spatial autocorrelation analyses is also discussed in §4.7.2.

3.6.5 Considerations in implementing spatial statistical infilling models

The value of spatial statistical models (§3.6.3–3.6.4) is that they permit a statistical evaluation of the process – to be able to say how good a technique is in terms of the strength and significance of relationships used to build the infilling model. The following considerations will help design a successful protocol, while recognizing model limitations:

Spatial heterogeneity. As noted earlier (§3.3.2: *Spatial consistency*), spatial predictive power drops off with distance as a function of climate variable, time aggregation interval, and heterogeneity in climate-controlling factors. The effect of climate heterogeneity on a spatial model is shown in Figure 11 – while this illustration is for a kriging model, the issue pertains just as well to simple regression schemes.

Because of heterogeneity in climate-controlling factors, the best predictor may not be the closest station. Consider a number of neighboring stations as candidates for simple regression models (as in the implementation strategy in §3.6.3). One approach is to stratify a regional domain by controlling factors (e.g., topography) and look for stations in a stratum matching the target station (this is the approach in PRISM: Daly et al. 2002, 2008).

Correlations dynamic. Spatial correlations among stations are dynamic, e.g., with strong seasonal dependence. Allow for station selection and regression models in your protocol to vary at least on a monthly basis or with a moving window.

Timestep choice. Infilling can be applied at different timesteps with varying results (Figure 13). As noted (§3.3.2: *Spatial consistency*), spatial correlations are smoother at longer timesteps. However, if, for example, daily missing values are relatively limited in number and scattered (rather than forming long gaps), a better monthly record may be obtained infilling at the daily scale because more of the original record is retained (see Figure 13 caption).

It is best not to use regressions based on monthlies to infill daily values – monthly regressions will not properly capture daily variability.

Station density effects. Station density has a direct effect on how well we can model spatial relationships. For the kriging model discussed in Figure 11, model error is larger during periods of low station density (Figure 12a, b). A consequence of this is that as station density drops off, interannual variability is artificially diminished. This is because with lower station density, the scheme reaches farther out to less well correlated stations, so that poorly related signals are blended to create the infilled record (Figure 12c vs. d). This variance loss can be accounted for by adding a stochastic term into the spatial model (Cressie 1993b).

---

[63] Following the notation in §3.4.4, the anomaly series $\delta\, x(t) = [x(t) - \bar{x}]$ and the standardized anomaly = $[\delta\, x(t) / SD_x]$.
[64] Advanced methods for spatial and *spatiotemporal* data are covered in Banerjee et al. (2003).

3.6.6 Alternate methods – When regression techniques do not work

Regressions can fail for stations where proximity and similarity in climate otherwise suggest they should work. Reasons for such failure include:

- Period of overlap between target and nearby station is too short on which to base a reliable regression relationship
- While both stations have comparable climate dynamics, daily events or even weekly values are not in phase over moderate spatial scales – e.g., under convective precipitation regimes.

The following alternate techniques, while practiced, are only advised in the case when statistical methods have failed. Generally speaking, this is because they are not statistically testable and because they can introduce undesirable features into data timeseries.

Longterm mean substitution. Insert the longterm (or some other period) day-of-the-year or monthly mean for missing days. Undesired effects are that (i) this technique under-represents temporal variability and (ii), for precipitation, it inserts spurious precipitation events into the record.[65]

Neighbor substitution. Insert the day's value from a neighboring station (or grid point) with offset or amplitude adjustment (§3.4.4). An advantage is that this process maintains some variability in the record based on the dynamics of a nearby station. A disadvantage is that there is no test as to whether the variability at the nearby station or grid point is related to that of the target station. In practice, this means that strong justification of station selection (or gridded dataset) is needed. Baron (2006) used this method to scale a gridded dataset's grid point timeseries to a station.

Distance-related interpolation. Use an inverse-distance (or other distance-dependent) interpolation of values from multiple neighboring stations (Wilmot et al. 1985, Chen et al. 2008). These spatial methods are distance-weighted in contrast to the previously presented statistical spatial models which are correlation-optimized. An undesirable effect is that this technique blends signals of the nearest stations – if these signals are not in phase, the result is a dampened signal.

A variant of the distance-based scheme is to interpolate station anomalies (as suggested for statistical spatial models, §3.6.4). One implementation is to use a variance-scaled anomaly (standardized anomaly):[63]

$$\delta(t) = \langle \text{INTERPOLATION}_{(i=1,n)} \{ [x(t,i) - \bar{x}(i)] / SD_x(i) \} \rangle \qquad (6)a$$

$$y'(t) = \bar{y} + [\, \delta(t) \times SD_y \,] \qquad b$$

where $\delta(t)$ is an interpolated standard deviation-normalized anomaly at time (t), derived from $x(t,i)$, $\bar{x}(i)$, and $SD_x(i)$ which are observations, mean, and standard deviation for each of the i=1,*n* number of *neighboring* stations used in the interpolation.[66] In eq (6b): $y'(t)$, $\bar{y}$, and SD_y are the target station's infilled value at time (t), mean, and standard deviation, respectively. This technique adjusts for signal mean level and variance difference not accounted for in the basic distance interpolation scheme; however, signals still have the potential for being smoothed.

---

[65] By using a mean value, an artificial, low magnitude 'event' is inserted in the record regardless of the synoptic weather conditions at the time – such as in the middle of a dry spell.

[66] The INTERPOLATION function can be any distance-related or other reasonable spatial interpolation scheme.

3.6.7 Temporal models – Serial correlation simulation

Data simulation based in temporal autocorrelation is an alternate, statistical technique. If a station's serial correlation is high, a temporal autoregressive (AR) model can be created to infill data using values in a station record on either side of missing points.[67] Regression terms in an AR model are observations at lag/lead times – most of the power in AR models tend to be at a lag/lead of 1 timestep so that a first-order model is sufficient. This approach can be valuable when there are no surrounding stations on which to base spatial methods as it uses only the target station's record. The disadvantage of this technique is that while it infills a record with an event structure that is characteristic of the station, the estimated values are synthetic (the values do not reflect actual events). Techniques that combine temporal and spatial correlations are mentioned at the end of §3.6.4.

More advanced versions of these models look for correlations with other variables at a station to guide this infilling. While used in other contexts, WGEN (Richardson 1981, Parlange and Katz 2000) is a temporal model that uses daily autocorrelation and cross-correlation[68] to stochastically simulate daily minimum and maximum temperatures and transition probabilities to model precipitation events.[69]

3.7 Document data changes and evaluate consequences

When you have a processed dataset ready to address your research questions, finalize the documentation of your data protocol. Layout your data clean-up methods, catalog dataset versions, record flag coding for adjusted, infilled, or omitted data (§3.2), and report tests that support your choice of methods. It is crucial to state assumptions, limitations, and caveats that accompany the techniques applied.[70] As mentioned earlier, keep in mind that estimated fields can in some cases artificially reduce variance in the data and lead to an overstatement of significance in statistical tests. Such frank discussion will be an important reference for others using the dataset, but also for you in interpreting your analyses. Do not neglect the documentation process.

Evaluate the effect of data adjustments and infilling as they carry into the statistical analysis phase (§4.0). Run different levels of unmodified and corrected data through to see how critical your decisions were, how robust your analysis results are. Be convinced that your data preprocessing choices had effects you are comfortable with. As part of documentation, do not neglect the evaluation process.

4.0 Analysis – Tools to Explore Critical Questions

With a credible, well-documented dataset tailored to planned analyses, you can pursue two complementary lines of study: discovery and hypothesis testing (Figure 1c). Discovery is aided by visualization and descriptive statistics, which help to develop an intuitive sense of the data, generate new hypotheses, and relate information to others. Hypothesis evaluation advances, with confidence, our understanding of a system and is accomplished through rigorous statistical tests.

---

[67] Wilks (2006: §8.3.1)[12] discusses temporal autoregression and AR models.
[68] Serial correlation (autocorrelation) is the lead-lag correlation for a single variable and cross-correlation lag-lead between variables.
[69] WGEN is designed to generate synthetic month-long and longer daily time series using a first-order Markov chain-exponential model for precipitation and a first-order autoregressive model for daily temperature that was conditional on precipitation. (Figure 18 shows the frequency distribution of WGEN-simulated dailies compared to station records)
[70] For one approach, see Kittel et al. (2004: §7.1)'s caveats.

Both discovery and hypothesis testing are components of analyses commonly applied to climate data, reviewed in the following sections. After discussing descriptive methods (§4.1), I introduce common derived variables that can help in such exploration (§4.2) and present temporal and spatial analytical techniques. Temporal analyses include those for:

- Longterm trends (§4.3)
- Covariation among timeseries (§4.4)
- Interannual variability and regime shifts (§4.5)
- Daily event structure and extremes (§4.6)

And spatial analyses for:

- Regional coherence (§4.7)
- Hemispheric teleconnections (§4.8)

I end with caveats re interpretation (§4.9) and a summary of key lessons for dataset creation and analysis (§5.0).

4.1 Description

Before proceeding with statistical tests, take advantage of descriptive methods to understand your data and visually explore your ideas. This process can reveal problems not already caught in dataset development (§3.0) and suggest the best analytical approach for testing your *a priori* hypothesized relationships (§2.0).

4.1.1 Descriptive statistics

Key descriptive statistics are:

- Data distribution parameters – e.g., mean, SD, quartiles, skewness, outliers
- For monitoring, diagnostic measures of recent observations relative to the historical record – e.g., departures from longterm mean, historical rank for extreme events. Departures can be expressed in original units, percentiles, or standard deviations.

4.1.2 Graphic methods

Helsel and Hirsch (2002: Chapters 1 and 2)[9] and Wilks (2006: Chapter 3)[12] present a wide variety of quantitative and graphic exploratory methods tailored to different discovery objectives. I list common graphic methods here, given by visualization goal and with examples from figures in upcoming sections (§4.3-4.8):

Data distribution (Univariate) –

- Histograms – to show the frequency distribution of a variable across:
  - Its own data range (e.g., Figure 17a,c)
  - Range of a second variable influencing the first's occurrence (e.g., by month: Figure 17b)
- Cumulative distribution graphs – to compare data distribution patterns of
  - Two datasets (Figure 20c)
  - A dataset vs. a hypothesized distribution
- Box-and-whisker plots (§4.6.1) –
  - To display data distribution features such as median, quartiles, outliers (Figure 18)

- Using multiple plots, to compare these features for different sites (Figure 18), seasons (Toews et al. 2007), or decades or by any other discriminating attribute influencing the variable's frequency distribution.

Multivariate: Relationships with other system variables –

- *x-y* scatter plots – to explore:
  - How two variables co-vary
  - How data from different categories of observations (different regions, seasons, etc.) break out in *x-y* space (e.g., with data domains delineated as in the style of Figure 2d).[71]

Temporal dynamics: To explore trends and interannual variability patterns –

- Timeseries plots – with values plotted on an absolute scale (Figure 15a) or as deviations from the longterm mean (Figure 15c), with or without smoothing (§4.5.1).
- Complemented by plots in the *frequency domain* (§4.4.1) to indicate periodic behavior (Figure 16a, Figure 22).

Spatial display: To explore geographic relationships –

- Mapped variable fields – a variable's distribution in space revealing gradients or distinct domains, such as with contoured fields (Figure 19c) or symbols (arrows in Figure 23a). Maps can be in absolute values or as anomalies from a spatial mean (e.g., contours and arrows in Figure 23a).
- Correlation maps – mapping temporal correlations of a spatially-distributed variable (such as temperature timeseries) with another, single timeseries.
  - The single series can be of the same variable at one key location – to illustrate spatial autocorrelation (§4.7.2; Figure 21a).
  - Alternatively, the single series can be of a second variable – to show their cross-correlation, such as with a hemispheric circulation index illustrating teleconnections (§4.8.2; Figure 24a).

Space-time: Evolution of spatial patterns with time –

- Animations – to display a time sequence of maps
- Time-longitude/Time-latitude section plots – two-dimensional display of spatiotemporal data, where one spatial dimension is collapsed by averaging (e.g., Figure 14).[72]

---

[71] See also "polar smooth plots" in Helsel and Hirsch (2002: §2.3.2, Figure 2.28).[9]

[72] In addition to animations, another means to display the evolution of spatial processes with time are *time-longitude* and *time-latitude section plots* (also known in meteorological applications as Hovmöller diagrams). In this method, three-dimensional spatiotemporal data – two dimensions in space (e.g., latitude by longitude) and the third in time – are represented in 2-D by collapsing one of the spatial dimensions into an average. These plots cross-section the data by time and one spatial dimension, and so reveal how a variable's spatial pattern evolves over time along a longitude or latitude transect:

- In a *time-longitude section plot*, averages over a latitudinal range are calculated by each longitude position and timestep, and then presented as a contour plot with longitude on the *x*-axis and time on the *y*. This shows how a variable generalized for a latitude range changes with time and longitude (e.g., Figure 14).
- In a *time-latitude plot*, longitude averages are contoured on a plot with time on the *x*-axis and latitude on the *y*.

In sector plots, spatial dimensions need not be latitude/longitude. A method is at http://locust.mmm.ucar.edu/episodes/episodes_paper_technote.html#_Toc516370405 – their technique is highly detailed, specific to their data, but the figures illustrate the process. For additional examples, see the interactive site for tropical climate variables: http://www.pmel.noaa.gov/tao/jsdisplay/.

4.2 Insights through derived variables

Analysis of composite and other derived variables can provide additional insights into site climates in terms of their thermal and drought regimes.

4.2.1 Thermal regime measures

Three commonly derived thermal variables are:

Diurnal temperature range (DTR). Diurnally, minimum and maximum temperatures (T_{min}, T_{max}) are strongly controlled by different local processes, those affecting daytime heating vs. nighttime cooling. However, day to day, they tend to be correlated because of multiday- to month-timescale effects – e.g., air mass advection and seasonal heat storage. This high correlation tends to obscure differences in T_{min} and T_{max} dynamics. On the other hand, DTR and mean temperature (T_{mean})[73] are generally orthogonal[74] and can be used to segregate processes controlling a climate's thermal regime – with T_{mean} showing multiday synoptic and seasonal effects and DTR the daily offsetting of local heating and cooling.

In terms of observed climatic trends, T_{min} and T_{max} have been changing at different rates over the recent record. We see that minima are often rising more strongly than maxima, so that DTR has narrowed with time (Easterling et al. 1997). DTR is a simple index of this change.

Accumulated growing degree days (AGDD). Accumulated growing degree days is a frequent measure of growing season conditions. By 'growing season,' we are generally referring to terrestrial plant phenology, but recognize that other groups of organisms have their own environmental cues. AGDD is the sum of mean daily temperatures that exceed a critical base temperature. Base temperatures (T_{base}) are set depending on application or ecosystem. A T_{base} of 5ºC has been used in general applications for temperate natural systems (e.g., Rehfeldt et al. 2006). However, where literature can support it, the limit should be one fitting the ecosystem or organisms studied. In montane and alpine environments, for example, 0ºC is considered a lower limit for plant growth (Billings and Bliss 1959, Kimball et al. 1973). The formula is:

$$AGDD = \Sigma[T_{mean}(t) - T_{base}], \text{ for days (t) when } T_{mean}(t) > T_{base}\ , \qquad (7)$$

summed over a year, and where $T_{mean}(t)$ is the average of daily minimum and maximum temperatures for day (t).

Frost-free period timing and length. The frost-free period is another indicator of growing season conditions. We can ask not only how the length of frost-free period changes year to year, but also about spring onset and fall termination date changes – as shifts in these two dates are not necessarily linked. Different ecosystems (and their different components) have different sensitivities to freezing temperatures – so we can use different cold temperature thresholds, T_{freeze}, depending on the application. A natural threshold for T_{freeze} is 0ºC, but a lower threshold such as –2 or –3ºC can be used to represent a 'hard' frost.[75] Example seasonal markers for spring and fall dates are:

---

[73] Both calculated from T_{min} and T_{max}: DTR = ($T_{max} - T_{min}$) and, of course, $T_{mean} = (T_{max} + T_{min})/2$. DTR was discussed earlier in the context correcting station records (§3.4.4: Considerations for specific variables).

[74] I.e., independent. This is generally speaking the case for T_{mean} and DTR at daily timescales; on the other hand, DTR can vary seasonally with T_{mean} and regionally, for example, maritime areas with low seasonality in T_{mean} tend also to have low DTR compared to continental climates.

[75] Note that many poikilotherms and plants (once hardening has begun) are not susceptible to freezing at temperatures a few degrees below 0ºC because of tissue solute levels and physiological adaptations (cf. Marchand 1996).

- Last and first frost at *night* – based on $T_{min} \leq T_{freeze}$
- Last and first day with freezing *daytime* temperatures – based on $T_{max} \leq T_{freeze}$
- Last and first *run* of, for example, 3 days with nighttime frost – based on $T_{min} \leq T_{freeze}$

4.2.2 Drought indices

The purpose of drought indices is to capture the occurrence and duration of wet and dry spells. Heim (2002) reviews commonly applied drought indices; two well-known ones are:

Palmer Drought Severity Index (PDSI). The Palmer Drought Severity Index is a standardized measure of soil moisture supply, typically evaluated on a monthly basis (Palmer 1965).[76] PDSI is a common metric for determining when a dry or wet spell begins and ends, integrating the effects of both precipitation and temperature (through its control over evaporative demand) on surface water balance (e.g., Figure 14). While reported for a given point in time, the index includes antecedent soil moisture conditions and so reflects the accumulative effects of water deficit or surplus. Alley (1984) presents a method for calculating monthly PDSI based on monthly precipitation and temperature data.

While PDSI's soil water budget allows it to integrate effects of temperature and precipitation and to accumulate moisture deficits or surpluses, key limitations come from the budget's shortcomings. These include (1) difficulty in applying it over terrain with heterogeneous soils and topography and (2) its lack of runoff generation lags, of snow and frozen-ground lag effects, and of seasonally in the role of vegetation (Alley 1984, Heim 2002). Relative to SPI (discussed next), PDSI may be slow to identify an emerging drought and may underrate the magnitude of prolonged drought (Karl and Knight 1985). Heim (2002) further explores the utility and limitations of PDSI.

Standardized Precipitation Index (SPI). The Standardized Precipitation Index focuses solely on the precipitation component of drought and wet periods (McKee et al. 1993, Guttman 1999). SPI is positive for wet conditions, negative for dry. To give temporal context to current drought or water surplus conditions, the index is determined for retrospective timescales from most immediate (proximate month and season) to sustained (multiyear) durations (Heim 2002).[77] SPI compares the current period's cumulative precipitation to the historical probability of reaching that amount of precipitation.[78] The SPI is the number of standard deviations that current precipitation totals are away from the historic median. The probabilities can be easily backed out of SPI using rules for normal distributions, e.g., ±2 SD correspond to roughly to the 2nd and 98th percentiles, respectively – so SPI values of –2 and +2 represent extreme dry and wet

[76] A global, coarse-resolution PDSI dataset is available at: http://www.cgd.ucar.edu/cas/catalog/climind/pdsi.html .

[77] Current SPI maps for the conterminous U.S. (by Climate Division) for 1- to 72-month durations are at: http://www.wrcc.dri.edu/spi/spi.html; see also: http://lwf.ncdc.noaa.gov/oa/climate/research/prelim/drought/spi.html. A historical, global gridded SPI data viewer is at: http://iridl.ldeo.columbia.edu/SOURCES/.IRI/.Analyses/.SPI/SOURCES/.IRI/.Analyses/.SPI/SOURCES/.IRI/.Analyses/.SPI/SOURCES/.IRI/.Analyses/.SPI/html+viewer? – zoomable, interactive.

[78] This probability is taken from the *cumulative distribution function* (CDF) for the location's precipitation record. This function is derived (accumulated) from a gamma *probability density function* (PDF) fitted to monthly precipitation data. (A gamma distribution is often a good portrayal of precipitation data.) The relationship between cumulative and probability density functions is illustrated in Figure 20a; see also §4.6.3. An inverse-normal transform of the gamma-derived CDF nicely puts these probabilities in terms of standard deviations of a normal distribution – the value of SPI is the number of standard deviations from the longterm median precipitation

conditions relative to the historical record for that site and timescale.[79] Guttmann (1999) lays out the calculation method.[80]

Benefits of SPI include: (1) that it is relatively straightforward to calculate (requiring only precipitation records and no site attributes), (2) its short to longterm perspectives on current drought conditions, and (3) advantages over PDSI noted earlier. However, a key limitation is that it only evaluates the role of precipitation anomalies in the occurrence and intensity of drought, while temperature anomalies can be an equally strong contributing factor (Hu and Willson 2000).

4.3 Trend analysis

4.3.1 Regression analysis

A method for statistical evaluation of trends in climate and other environmental variables is regression analysis.[8] Regression trend analysis applies eq (1) with x(t) now representing time t itself:

$$y'(t) = b_0 + b_1\, t \tag{8}$$

where the regression slope, b_1, is the calculated trend and y'(t) traces the variable's change due to the longterm trend alone.

Regression model assumptions. While regression is frequently used to study trends, it is not always an appropriate method. Referring back to earlier discussion re regression (§3.4.3.1), the standard implementation (OLS), while powerful, requires that observed data are independent (discussed shortly) and their distribution normal and homoscedastic (variance constant).[38] As noted previously, climate data do not always meet these criteria. Transformations to adjust data to normality and constant variance are noted in §3.4.3.1, along with corresponding issues.

If the two distribution assumptions are violated (even after transformation), use a robust technique –

- If non-normal – use a non-parametric method, such as Mann-Kendall test for trends (next section, §4.3.2)
- If normal, yet heteroscedastic – use a robust regression which models the changes in variance, such as WLS regression (§3.4.3.1)[40] and quantile regression (§4.3.3).

Adjusting for serial correlation. In timeseries, independence means that the observations are not serially correlated (no temporal autocorrelation). However, it is not unusual for geophysical data to be serially correlated. Needless to say, these assumptions are often ignored and regression trend statistics commonly reported in studies. A test for serial correlation is the Durbin-Watson test, which is applied to the detrended series[81] (Wilks 2006: p. 192; Helsel and Hirsch 2002: §9.5.4.1).

---

[79] As noted in footnote 78, the probabilities behind the SPI are transformed to a normal distribution. Conversions from standard deviations (and so, SPI) to percentiles are based on one-tailed probabilities. Other rough conversions are: ±1 SD = 16th and 84th percentiles, ±1.6 SD = 5th and 95th percentiles, ±2.3 SD = 1st and 99th percentiles, ±3 SD = 0.1th and 99.9th percentiles (see also Guttmann 1999). Note that distribution function fitting does not always do well at extreme tails of a distribution, so there is high uncertainty in extreme SPI values. (See related discussion: http://www.cpc.noaa.gov/pacdir/NFORdir/INTR.html – "Caution Required for the Tails of the Curves".)

[80] PACN (2008) gives a procedure for calculating SPI in Excel. Inclusion here is not an endorsement – before using, validate output against results from other methods.

[81] A detrended timeseries is usually calculated as the residuals from linear regression (trends) analysis.

If serial correlation is significant and the correlation is positive at a timestep lag of minus 1 (r_1), then its effect can be accounted for in trend analysis by inflating the regression's mean sum of squares for error (*MSE*). This appropriately makes the regression's significance test more conservative. The inflation factor increases with r_1, such that:

$$MSE^* = MSE\ [(1 + r_1) / (1 - r_1)]\ , \qquad (9)$$

where *MSE** is the serial-correlation adjusted value of *MSE* (Wilks 2006: p. 194). From the adjusted MSE*, recalculate R^2, *F*-test statistic, and significance level *p*. [82]

4.3.2 Mann-Kendall trend test

The Mann-Kendall test for trends is a non-parametric method using ranked data. Advantages of this test are that it:

- Does not require normally-distributed observations
- Allows for missing data
- Is resistant to the effect of outliers
- Detects monotonic rather than strictly linear trends – with the added benefit of eliminating the need for nonlinear transforms
- Is often just as powerful as corresponding parametric tests such as regression (Lettenmaier et al. 1994)

While normality is not needed, constant spread in the data's distribution (homoscedasticity) is. However, with this relaxed set of requirements, workable transformations are more readily achieved than under the more restrictive assumptions of ordinary least squares regression.[83] In the Mann-Kendall test, climate observations are converted to ranks and the ranked correlation (climate rank vs. time) is tested with Kendall's *tau* statistic (Helsel and Hirsch 2002: §8.2, §12.2.1).[9] Hirsch and Slack (1984) give a method accounting for seasonal dependence in trends.

Because observations are ranked, this technique tests for monotonic trends (as already noted) – with the advantage of not needing to specify a linear or nonlinear model,[84] but with a notable disadvantage of not quantifying that trend. Hirsch et al. (1982) present a slope estimator, Kendall-Theil robust line, to accompany Mann-Kendall tests.[85] An alternative strategy is to use (1) the Mann-Kendall test for statistical assessment and (2) linear regression (with or without nonlinear

---

[82] To calculate the serial-correlation adjusted R^2* and *F**:
- Back out the error sum of squares (SSE) and total sum of squares (SST) from adjusted MSE* using the definitions:
  - MSE* = SSE*/(*n*-2) → SSE* = MSE*×(*n*-2)
  - SST* = SSR + SSE*

  where * indicates an adjusted parameter, *n* = number of observations, and SSR is the sum of squares for regression obtained from the un-adjusted regression analysis.
- Recalculate R^2* and *F**, get new *p*
  - R^2* = 1 – (SSE*/SST*)
  - *F** = MSR/MSE*
  - Determine the *p* from *F** and degrees of freedom for MSR (*df*=1) and MSE (*df*=*n*-2) [or compare *F** to $F_{critical}(\alpha)$]

  where MSR is the mean sum of squares for regression also obtained from the un-adjusted regression analysis.

[83] If corrected by a power transform, the Mann-Kendall test statistic *tau* (discussed next) is variable-scale independent – that is, not affected by the transform and so is comparable across tests on original and variously transformed data (Helsel and Hirsch 2002: §12.2.1)[9] in contrast to scale-dependence in R^2 (§3.4.3.3: *Transformations and* R^2).

[84] vs. for regression techniques, §3.4.3.1

[85] See Helsel and Hirsch (2002: §10.1)[9] re the Kendall-Theil robust line

transformations) to describe the trend.[86,87]

Accounting for serial correlation. A concern in applying Mann-Kendall is, like parametric tests, that it assumes independent observations. In evaluating this issue, Harcum et al. (1992) found that the test considerably overstated a trend's significance level when the detrended serial correlation at 1-year lag had a correlation coefficient $r_1 > +0.2$. They concluded that for $r_1 < +0.1$, the test was rigorous, with a grey area between +0.1 and +0.2.([88]) Evaluate this issue by (1) detrending the series[81] and (2) calculating the lag-correlation between this series and itself offset by one timestep. Kulkarni and von Stroch (1995) and Hamed and Rao (1998) offer methods for handling serial correlation in the Mann-Kendall test.

4.3.3 Quantile regression – Tracking heteroscedasticity

So what if a process is heteroscedastic with time – that the variance and other moments have trends? If we can see how a process varies over time in a manner far more complicated than just a changing mean, we gain deeper insights into how the system is actually working (Cade and Noon 2003, Beniston and Stephenson 2004). While standard regression considers the tendency of the mean, *quantile regression* follows the trends of quantiles (percentiles) of a variable's distribution evaluated for intervals along the time axis. It allows us to simultaneously track any part of the distribution, such as the median (50%-tile) and top vs. bottom 10%-tiles. The approach is robust as it models heteroscedasticity and has low sensitivity to outliers (Hao and Naiman 2007). Cade and Noon (2003) present examples of quantile regression from ecology. General references are Koenker (2005) and Hao and Naiman (2007).[89]

4.4 Covariation among variables

4.4.1 Relationships in time and frequency domains

If two variables both vary in time, we can ask whether they co-vary in a manner that implies cause and effect (regression) or coordination (correlation). In park applications, these variables may be both climatic or one that is resource related, following research questions. Parametric and non-parametric regression methods just noted for trend analysis, and corresponding correlation techniques, can similarly be employed to compare two such timeseries: x(t) vs. y(t). For example:

- For regression –
  - OLS and WLS regression (§3.4.3.1, §4.3.1)
  - Kendall-Theil robust line (§4.3.2)
  - Quantile regression (§4.3.3)
- For correlation[90] –
  - Pearson's r
  - Rank correlation methods: Kendall's *tau*, Spearman's *rho*

---

[86] In this approach, it is important to distinguish the roles of these two analyses when reporting results. While the Mann-Kendall test evaluates the statistical significance of an observed trend, it does not similarly assess the regression's quantification of the trend.

[87] Helsel and Hirsch (2002: §10.1.2)[9] compare Kendall-Theil robust line and ordinary least squares regression slope approaches.

[88] For the mirrored case where $r_1 < -0.2$, autocorrelation leads to an understatement of statistical significance (Hamed and Rao 1998).

[89] Koenker created an R-package for quantile regression: http://cran.r-project.org/web/packages/quantreg/index.html (http://cran.r-project.org/web/packages/quantreg/vignettes/rq.pdf). Koenker (2005) and Hao and Naiman (2007) are available in part on Google Books (see References).

[90] See Helsel and Hirsch (2002: Chapter 8 – Correlation)[9]

Processes that vary in time lend themselves to additional questions and corresponding techniques:

- *Cross-correlation* and *Multi-lag regression* – To test for a *lag* or *lead* of one variable's dynamics relative to the other.[68,91,92]
- *Cross-spectra* – In addition to evaluating relationships in the *time domain* (as in methods just covered), we can ask how correlation is structured in the *frequency domain*.[93] Specifically, is their correlation (*spectral coherence*) concentrated in certain frequency ranges? And are their dynamics in these frequencies in *phase*, or do they lag/lead? Cross-spectral techniques are presented in §4.7.3, in the context of comparing station records of the same variable.

4.4.2 Cautions regarding timeseries comparisons

In addition to taking care to adhere to these tests' assumptions (re data distribution, serial correlation, etc.) and regression implementation and interpretation guidelines presented earlier (§3.4.3), take note of the following cautions for timeseries comparisons:

Smoothed dataseries. Smoothing timeseries underrepresents observed variance (lost in temporal averaging) and increases serial correlation (information is blended among adjacent time points). Both effects compromise statistical analysis – they increase the risk of overstating significance in statistical tests (Type I error).[24] Consequently, smoothed data should be strictly avoided in statistical comparisons (discussed further in §4.5.1: Additional cautions).

If timesteps do not match. Trying to compare timeseries with differing timesteps can be troublesome. Some points:

- The high-frequency variance of a short-stepped timeseries is missing in long timestep data, and so cannot contribute to shared variance in correlation or explained variance in regression. Instead, this variance ends up in the error term, reducing the power of these tests.
- There are various approaches to force timesteps to match. If the longer timestep series is interpolated to the shorter step, the resulting series becomes strongly serially correlated (information is repeated at the more frequent time points) and the apparent number of observations (and so degrees of freedom) becomes inflated. If unaccounted for, these effects increase the likelihood of a Type I error (overstating significance).
- The more common approach is to rescale short timestep data to the longer step, such as through aggregation as by summing or averaging across fixed time-intervals[94] (giving, for example: AGDD, total annual precipitation, and mean monthly temperature) or binning by event categories (e.g., number of extreme cold events in winter).[95] If rescaling is done

---

[91] A couple examples of studies using lag-lead cross-correlation between climate and biological variables are: Martinez-Yrizar and Sarukhan (1990), Braswell et al. (1997).

[92] Serial correlation in individual timeseries also interferes with cross-correlation tests. As an alternate approach to cross-correlation, Burnaby (1953) presented a test for comparing timeseries that are autocorrelated. Malmgren et al. (1998) used this technique to evaluate teleconnections.

[93] In the time domain, variables are paired by time index: $y(t)$ vs. $x(t)$. In the frequency domain, they are indexed along a range of possible oscillation frequencies, f: $y(f)$ vs. $x(f)$. In spectral techniques, f is generally treated as a narrow frequency band, rather than a single frequency. Univariate spectral analysis is presented in §4.5.2, bivariate (cross-spectral) analysis in §4.7.3.

[94] As distinguished from running averages, which constitutes smoothing.

[95] As another example, the Climate Extremes Index (Gleason et al. 2008) uses a season-aggregate index of extreme events (Figure 15a).

through smoothing, however, issues discussed above arise. With aggregation, keep in mind there is a loss of shorter-term information – a concern if scale interaction plays a role in the processes being compared (a topic covered shortly: *Process timescales differ*).

Processes out of phase. If one process has a lagged effect on the other, standard correlation and regression techniques are likely to miss or underestimate a relationship. Cross-correlation and multi-lag regression (as just noted, §4.4.1) can be used to explore this possibility.

Process timescales differ. Regression and correlation analyses are straightforward when the temporal scales of processes of interest reasonably match.[96] On the other hand, in looking for relationships between a slow-moving process and a fast one, the mismatch in scale complicates matters. This is because (1) processes may only interact in a narrow, shared range of frequencies or (2) they interact across scales (§2.2: *Scale interactions*). The three modes by which slow and fast-moving processes interact and corresponding analysis options are:

- *Mode 1: Processes interact only at shared frequencies.* Cross-spectral analysis (§4.7.3) can be used to show at what frequencies two processes correlate well and if this coherence is in phase or lagged. Note that coherence and phase do not prove there's a physical link between the two processes, but can be the basis for hypothesizing mechanisms.
- *Mode 2: Slow process constrains a fast process*. From the perspective of a fast process, a slow process operates between relatively stable states (phases) and potentially constrains the dynamics of the fast process. We can ask if these phases set the stage for the high-frequency process, resulting in its distinct behavior. An analysis strategy to assess such phase dependence is to:
  (1) Block the slow-moving dataseries into periods when the system is in more or less stable phases.[97]
  (2) Contrast behavior of the fast process under the different blocks of the slow process. This can be evaluated using analysis of variance (ANOVA) or other parametric and non-parametric techniques to test whether a significant part of faster process's variance is explained by the slow process's phases.[98]
- *Mode 3: Fast process determines the state of a slow process.* Critical, threshold events in a fast process (e.g., extreme weather events) may control the outcome of a longer-acting process. The trick in this type of analysis is to select the appropriate fine-scale variable, such as crisis weather events – this selection is aided by having detailed knowledge of the system. For example, in an ungulate demographic study, Hallett et al. (2004) evaluated the effects of irregularly-timed winter weekly low temperature, high rainfall, and high wind events on interannual variability in mortality.[99]

These approaches will only work, of course, if processes of interest are represented in the data at timesteps corresponding to the timescales at which they constrain or impact each other.

Interpretation of correlation and regression results. Caution is needed in interpreting both significant and nonsignificant results in correlation and regression analyses. As we are always

[96] Provided, needless to say, that variation at these scales are adequately captured by their datasets.
[97] Data blocking is also discussed in §4.8.2, re testing teleconnections. See also Helsel and Hirsch (2002: §7.3 – Blocking).[9]
[98] The two-sample *t*-test is another parametric technique for comparing means of one variable by blocks of another in the case of when there are just two blocks. Parametric and corresponding non-parametric tests are discussed by Helsel and Hirsch (2002: Chapter 7).[9]
[99] This study was also discussed in §2.2: *Scale interactions*

reminded, "correlation does not imply causation" – that is, be careful not to take significant results as validation of hypothesized mechanisms. See §4.9 for further discussion, especially in regards to nonsignificant results.

4.5 Interannual variability – Spectral analysis and regime shifts

An important temporal feature of climate is oscillatory patterns at interannual, multidecadal, and longer scales.[100] These often suggest a link to climate system processes that exhibit similar characteristic modes of behavior. Such oscillations tend to be quasi-periodic, that is, tending to fluctuate within a band of frequencies rather than with a strict return period. Their dynamics are often partly obscured by 'noise' and processes operating at other frequencies and so need to be examined using techniques that separate out signals by frequency bands. Signals of interest can be explored visually using low-pass filters (next section, §4.5.1) and evaluated numerically using spectral analysis (§4.5.2). Another key dynamic of climate is regime shifts at multidecadal and centennial timescales; I cover techniques for their analysis in §4.5.3.

Linkage of such local interannual variability to regional and global dynamics is explored later on in sections on spatial pattern analysis, §4.7 and §4.8, respectively.

4.5.1 Smoothing filters

Smoothing filters are low-pass filters, allowing only lower frequencies of a timeseries through. The simplest are moving averages (running means) using *unweighted averaging* (also, called 'rectangular filters') (e.g., Figure 15a). These are typically applied over an odd number of years and time registered on the middle year of the moving window.[101] Other common low-pass filters use *weighted averages* and span a sufficient number of years to remove high frequency variation in the data (e.g., Figure 15b, c).

Example weighted filters across a range of filter widths, as applied to annual data, are:

- 3-year, (1-2-1) weighted filter – this simple 'triangular filter' is commonly used for removing highest (interannual) frequencies. The weight denominator is 4, so the weights are ¼ - ½ - ¼.
- 5-year, (1-3-4-3-1) filter – for removing interannual through 3-5 year variability (cf. Trenberth et al. 2007: Appendix 3.A). Weight denominator = 12.
- More complex functions for emphasizing decadal processes – such as, the 13-year scheme in Figure 15b and applied in Figure 15c.

Numerical issues with smoothing include two artifacts:

End effects. Because these filters are calculated at the center point of the moving window, they cannot be calculated for the first and last few points of a series. The resulting timeseries will then be shorter than the original. These points can be left blank or padded with a value that makes sense in the context to your application: such as the longterm mean (zero, if an anomaly series) or the proximate filtered value.

---

[100] A primer on statistics related to climate variability is: http://www.nws.noaa.gov/om/csd/pds/PCU2/statistics/Stats/part1/SPrimer1.htm and .../part2/SPrimer2.htm. For accompanying glossaries for statistical terms: http://www.nws.noaa.gov/om/csd/pds/PCU2/statistics/glossary.htm and dynamical meteorology: http://www.nws.noaa.gov/om/csd/pds/PCU2/meteorology/glossary.htm

[101] Avoid using non-centered (e.g., prior-moving) averages as this offsets peaks and troughs with respect to actual timing. These are occasionally found in the literature and are what is provided by the moving average option in Excel's graph trendline function.

However, it may be desirable for these values to reflect information from the original series's end points. One method is to apply a filter with stepwise decreasing span length as the filter center approaches an end point – with the caveat that the ends will have higher frequency variability than in the main body of the smoothed series. Mann (2004) presents an alternate technique that optimizes a choice among three constraints on the smoothed line as it approaches the end points: these constraints are on the smoothed line's local departure from the longterm mean, slope steepness, and slope change (the second of these was applied in Figure 15c).[102]

Spectral side lobes. Some filters, including weighted filters, create 'side lobes' in power spectra (§4.5.2) outside of the frequency range intended to be captured. As a result, higher frequency variance inadvertently shows up in smoothed (low-pass filtered) signals.

Additional cautions are:

Not for statistical analysis. Smoothing is an exploratory tool, to aid visualizing temporal patterns. As mentioned earlier (§4.4.2), the resulting series has severely reduced variance and elevated serial correlation and so should not be used in statistical testing. They also should not be used in spectral analysis (§4.5.2) due to the introduction of side lobes and any end-effect corrections.

Keep these caveats in mind when obtaining processed dataseries from other sources, such as teleconnection indices (§4.8.2). To this end, check documentation to see if processing included smoothing and, if so, check for the availability of unsmoothed versions.

Missing the details. Smoothing climate data may become so routine for monitoring mid- and longterm behavior that we miss or discount unusual, yet critical events. As an example, Verosub and Lippman (2008) note this caution with respect to tracking effects of single-year global climatic events, such as volcanic eruptions, in regional and local climate records.

4.5.2 Spectral analysis – A look in the frequency domain

Spectral analysis is a method for formally exploring time series for oscillatory behavior (Yiou et al. 1996, Ghil et al. 2002). A common spectral technique is Fourier analysis which identifies at what frequencies the data most strongly vary. Results are commonly portrayed as a *power spectral density function*, where oscillation strength (spectral power) is plotted against frequency (Figure 16a). Peaks show at what periods (= 1/frequency) a climate record varies.

Narrower peaks reflect oscillations with tighter return periods vs. quasi-periodic dynamics shown by broader peaks. Occurrence of certain characteristic patterns can suggest linkage to continental and hemispheric climate dynamics, such as El Niño (more on teleconnections follow in §4.8). Evaluating whether spectrum peaks are statistically significant is an important component of spectral analysis (Figure 16a).[103] Without statistical evaluation, it's too easy to place undue significance on frequencies corresponding to peaks in spectral power.

---

[102] Mann's Matlab routine is provided at: http://www.meteo.psu.edu/~mann/Mann/tools/tools.html (relocated from that given in Mann 2004).

[103] Spectral analysis software with confidence interval capabilities include –

- IDL: Coherence function
- on-line spectral calculator: *Spectral Analysis (v1.0.6) in* Wessa (2009) http://www.wessa.net/rwasp_spectrum.wasp
- *Singular Spectrum Analysis - MultiTaper Method (SSA-MTM) Toolkit* (http://www.atmos.ucla.edu/tcd/ssa/)

These packages are presented as examples, not as an endorsement or reflecting an assessment.

Data are commonly preprocessed for spectral analysis by removing trends and, if present, the seasonal cycle. This processing removes corresponding peaks in the spectrum, making results more straightforward. Trends can be removed by determining residuals from linear regression;[81] the seasonal cycle removed by subtracting corresponding longterm seasonal means.[104] As discussed earlier (§4.5.1), spectral analysis should not be run on smoothed data because of:

- Filter side lobes – frequency spillover into higher frequency bands
- End-effect corrections, if incorporated into smoothed series – also blurring filter frequency boundaries

– both of which significantly affect spectral density (von Storch and Zwiers 2001).

The utility of Fourier-based spectral analysis is limited (1) when periodic signals are nonstationary (their frequency changes) and (2) for short timeseries (Kestin et al. 1998). Wavelet analyses is an advanced technique for following changes in the frequency of oscillations and will work for short records. Torrence and Compo (1998), Gedalof and Smith (2001), Ghil et al. (2002), Chang et al. (2004), Gray et al. (2003), and Labat (2005) present this method as applied to climate studies.

4.5.3 Regime shifts

Shifts in features that make up regional climates are not uncommon, having occurred at times over the last century and longer timescales (e.g., Pederson et al. 2006, Diaz et al. 2008) (Figure 16b). These shifts, which run through both the abiotic and biotic environment, tend to be linked to regime changes in hemispheric atmosphere and ocean circulation (e.g., Hare and Mantua 2000, Gedalof and Smith 2001; such cross-scale geographic linkages are discussed in §4.8). Regime shift step-detection techniques are presented by Box and Tiao (1975),[105] Biondi et al. (2002), and Rodionov and Overland (2005).[106]

Changes in regime are also expressed in records as breaks in trend slopes. Breakpoint analysis techniques in climate and ecological studies include piecewise-linear (segmented) regression (Tomé and Miranda 2004, Marlon et al. 2009)[107] and quantile regression (Koenker and Schorfheide 1994)

Regime analysis is best done on records that are from the start free of artificial temporal inhomogeneities (e.g., station changes) as these change points complicate shift detection. Note also that methods correcting data heterogeneities (§3.4) run the risk of removing true regime shifts. Such corrections should be (1) omitted from datasets bound for regime change studies or (2) applied carefully only for the clearest station change cases, and then well documented and followed up with a review of regime shift results for interference at correction change points.

4.6 Daily analysis – Structure and extremes

At finer temporal scales, analysis of daily and hourly records can reveal the characteristic structure of weather events and the frequency of extremes for a site. These features are generally a reflection of a region's climate. Shifts in their structure can reveal important climate changes as much as

---

[104] Longterm seasonal means are longterm means by day, week, month, or season corresponding to the timestep of the observed data. 'Longterm' averaging is over station record, 30-year normals, or other set period.

[105] See, for example, Pederson et al.'s (2006) implementation of Box and Tiao's (1975) intervention analysis and other techniques in the detection of regime shifts.

[106] Rodionov and Overland's model is available at: http://www.beringclimate.noaa.gov/regimes/index.html

[107] Tomé and Miranda (2004)'s method is implemented in Miranda and Tomé (2009). Marlon et al. (2009) used the R package 'Segmented' (Muggeo 2009). See also references re use of breakpoint analysis in testing record homogeneity (footnote 57).

altered means and interannual variability do. Meehl et al. (2000) and Trenberth et al. (2003) review modes by which the character of weather events change under altered climate. Among recent observed shifts in daily structure are more days with warm temperature extremes, fewer cold extremes, and greater occurrence of extreme daily precipitation (Easterling et al. 2000, Trenberth et al. 2007).

This section presents means for characterizing events (§4.6.1) and ways to assess their change in terms of:

- Event structure parameters (§4.6.2)
- Frequency distribution functions (§4.6.3)
- Extreme value analysis (§4.6.4)

Records with corrected or infilled values (§3.6) should be omitted from these analyses.

4.6.1 Event structure characterization and display

Event structure can be evaluated graphically with frequency distribution plots (Figure 17). In this technique, occurrences of a given event are 'binned' by:

- Event magnitude (Figure 17a, Figure 20d)
- Timing (season, time of day; Figure 17b)
- Duration or prevalence (Figure 17c, Figure 19b)
- Other variable offering a perspective on event structure.

A more analytical presentation is box-and-whisker or notch plots[108] which express the median, standard deviations (SD), quartiles, range of values, and outliers (Wilks 2006: §2.5)[12] (Figure 18). For variables with highly skewed frequency distributions (such as precipitation), values can first be transformed to normalize the data. As noted earlier, typical transformations for precipitation include natural log (Figure 18), square root, and cubic root (see §3.4.3.1 for methods; cautions re frequency-distribution transformations are covered shortly). Toews et al. (2007) present a box plot application for seasonal analysis of event structure.[109]

For precipitation and other discrete-event variables that have a reasonable likelihood of zero values (e.g., dry days), only non-zero values should be incorporated in box plots and related analyses.[110] These plots then show the structure of precipitation events just for when they occur. An additional graphic can be employed to show the frequency of whether or not there's an event (as in, number of wet vs. dry days).

Issues in the analysis of daily event structure include:

Observation biases. For precipitation, observer practices can lead to underreporting the smallest events and a biasing toward certain frequency bins over others (e.g., favoring recording daily values in multiples of 0.05 inches). Daly et al. (2007) evaluates the consequences of these biases.

Corrected data. Omit inserted precipitation daily values parsed from multiday accumulated totals (§3.3.3), infilled (§3.6), and other corrected data that may not adequately capture (and

[108] Box-and-whisker plots are also known as 'box plots.' The difference between box and notch plots is described in Figure 18.
[109] This application is for the 'R' statistical computing environment. For box plot graphing in general, PACN (2008) gives a procedure for Excel.
[110] Using only non-zero values is also practical if a logarithmic transform is selected [because *log*(0) is undefined].

instead blurs) a site's true daily event structure. Be aware that frequency distribution parameters and estimation of frequency distributions (especially with respect to extremes, §4.6.4) can be highly sensitive to data problems and their correction.

Back transformation of mean, percentile, and interval parameters. When nonlinear transforms are applied to event data, caution is needed in back transforming structural parameters to the original linear scale. The mean calculated in transformed space generally has little correspondence to the variable's actual mean (in linear space), while the median (50%-tile) and other percentiles do.[111] Any *interval* – such as interquartile range (IQR, cf. Figure 18), standard deviation (SD), and confidence intervals (CI) – determined in transform space and back transformed also has little meaning. Instead, an interval should be represented in terms of its *upper and lower limits,* where these are determined in transform space and back transformed to linear values.

4.6.2 Changes in event structure parameters

Altered daily or hourly structure can be assessed by analyzing changes in *event structure* parameters such as those used to characterize events (§4.6.1). These can evaluated simultaneously using quantile regression (e.g., Beniston and Stephenson 2004; §4.3.3) or individually with trend and interannual variability analyses (e.g., Figure 19d; §4.3, 4.5). Candidate parameters include:

- Frequency of events in a specific range of a binning variable – such as a specific bin in a frequency distribution histogram (as those in Figure 17; see §4.6.1 re binning)
- Frequency of events beyond a given threshold (threshold exceedance) – as might define extreme events and be based on, for example:
  - A threshold critical to the system – such as temperatures below freezing or a critical precipitation rate (e.g., Mearns et al. 1984, Beniston and Stephenson 2004)
  - An upper or lower percentile (e.g., Figure 19a; Climate Extremes Index thresholds per Gleason et al. 2008: Figure 15a)[112]
- Frequency distribution parameters – as those given in box plots
  - For example, the interquartile range (IQR, Figure 18) determined over short intervals in the record (e.g., by year or decade)

4.6.3 Changes in frequency distribution functions

Figure 19a shows changes in nighttime temperature extremes as a timeseries. Such change can also be evaluated in terms of shifts in its *relative frequency distribution* (Figure 19b).[113] Changes in this distribution can be analyzed using a test for identical distributions, such as the two-sample Kolmogorov-Smirnov test (Conover 1999).[114,115] This nonparametric test is sensitive to differences in shape or position of a frequency distribution – both types of changes are illustrated in Figure 19b.

To apply the Kolmogorov-Smirnov identical distribution test, we use a cumulative form of the relative frequency distribution of a variable. This is as the *cumulative probability distribution*

---

[111] Helsel and Hirsch (2002: §9.6.3)[9] discuss corrections for transformation biases in the mean.
[112] This would be a reduced form of quantile regression.
[113] Relative frequency distribution also referred to as the *probability distribution (or density) function* (PDF).
[114] The 2-sample Kolmogorov-Smirnov test for identical distributions is also known as the Smirnov test (Conover 1999). This test is distinguished from a *one-sample* Kolmogorov-Smirnov test, which is a goodness-of-fit test used to evaluate an observed distribution against a theoretical one (e.g., to test for normality[38]).
[115] An on-line calculator for the 2-sample test is at: http://www.physics.csbsju.edu/stats/KS-test.html (Kirkman 1996).

function,[116] where relative frequencies are accumulated (summed) along the *x*-axis until all observations are accounted for.[117] Figure 20a, b illustrate the relationship between these two distribution functions and how changes in the shape of one is reflected in the other. The Kolmogorov-Smirnov test evaluates if the separation (*D* in Figure 20c) between the cumulative distribution functions of two sets of observations is significant (Figure 20b, c, d).

Another technique for comparing frequency distributions is quantile-comparison (or quantile-quantile) plots, where the percentiles of two distributions are plotted against each other.[118] These graphics are most useful if accompanied by plotted confidence intervals or statistical tests to evaluate if the distributions are identical.

4.6.4 Extreme value analysis

Changes in occurrence of extremes can arise from (1) shifts in location, dispersion, and asymmetry of the frequency distribution (i.e., in the mean, variance, skewness, and other moments) and (2) changes in the structure of the far ends of the distribution (change in tail shape) (Meehl et al. 2000). While testing for overall changes in frequency distribution can rely on methods discussed in §4.6.3, evaluating trends in extreme events requires detection of changes in the tails of frequency distributions – where extremes lie but whose distribution is difficult to estimate by standard statistical methods.

Katz and Brown (1992), Kharin and Zwiers (2000), and Goubanova and Lia (2007) describe and implement an extreme value analysis technique based on a *generalized extreme value distribution* (GEV) which selects from three possible asymptotic distribution models fitted to extremes in the record. Frei and Schär (2001) present a method for detecting trends in extreme event return periods. Coles (2001) and Hosking and Wallis (1997)[119] provide introductions to extremes value analysis in theory and practice.[120,121]

4.7 Spatial pattern analysis. I: Regional connections

Once the temporal dynamics of a site's climate has been explored (§4.3–4.6), we can ask how its behavior fits in with the region or farther afield: Is the station representative of a park and the region? Or does its dynamics contrast with neighboring stations? Tools to assess spatial coherence include geostatistical and cross-spectral techniques (§4.7.1–4.7.3).

---

[116] Also referred to as the *cumulative distribution function* (CDF) or *cumulative probability distribution* (CPD)

[117] A re-expression of the cumulative distribution function is the *probability of exceedance* = [1 - (cumulative distribution function)]. The CDF gives the percentile for a value among all observations, the probability of exceedance gives the chance of observing a value (e.g., a precipitation amount) above a certain level and corresponding return period. This approach is often used in extreme value analysis techniques (§4.6.4) and extended-range weather outlooks: http://www.cpc.noaa.gov/products/predictions/90day/.

[118] Q-Q plots in Helsel and Hirsch (2002: §2.2.5)[9] and Wilks (2006: §4.5.2).

[119] Hosking and Wallis (1997) available in part on Google Books (see References).

[120] Online resources re extreme value analysis (see Stephenson and Gilleland 2005 for a review of software):

- http://www.cru.uea.ac.uk/projects/mice/html/extremes.html
- http://www.met.rdg.ac.uk/~han/Extremes/extreme1.pdf - a presentation (Stephenson 2002)
- For S-Plus, R, MATLAB, others: http://www.rap.ucar.edu/staff/ericg/softextreme.php; specifically for R: http://www.isse.ucar.edu/extremevalues/evtk.html. See also: http://www.rap.ucar.edu/staff/ericg/extremereading.html
- Statistical Tool for Extreme Climate Analysis (STECA): http://www.cics.uvic.ca/scenarios/index.cgi?Other_Data#steca

[121] See also Makkonen (2008) for additional discussion of this method and related problems.

These analyses can be complemented by on-line national monitoring and outlook products which provide a near-real time, regional perspective on a park unit's climate. A selection of these is presented in Table 2.

4.7.1 One-point correlation map

A *point-correlation map* displays the spatial distribution of correlation coefficients between the record at a single location with those of other points within a domain (e.g., Figure 21a). These maps are a straightforward means to reveal spatial connections between a site and surrounding region.[122] Interpretation of these maps is facilitated if statistical significance of correlations is also indicated (e.g., by stippling, as on Figure 19d in another application).

4.7.2 Spatial variation – Semivariograms

We discussed spatial autocorrelation earlier in the context of infilling missing data (§3.6.4). These techniques can also be used to assess how homogeneous vs. heterogeneous climate is across a domain. Related to autocorrelation, the semivariogram is the spatial variance as a function of distance.[123] The semivariogram reveals how rapidly or slowly stations become less related to each other the farther apart they are (Figure 21c).[124] The relationship with distance can additionally depend on direction (anisotropy; evident in Figure 21d – see caption discussion).

The semivariogram is usually assessed for a point in time in the record (e.g., a day, week, or month depending on the climate process being evaluated: e.g., February 1996 in Figure 21c). The shape of a domain's semivariogram often changes with season, linked to seasonal climate processes. Controlling for season, the semivariogram can be relatively stable over decades, especially if it is controlled by topography (Fuentes et al. 2006). At the event level, the shape can be tied to synoptic conditions (rainfall type) as well as elevation (Şen and Habib 2000). Mapping the semivariogram at a set distance (e.g., 40km in Figure 21d) can reveal a regional climate's spatial connectivity and anisotropy (Figure 21d, e – see caption). Spatial statistical references, techniques, and issues are discussed further in §3.6.4–3.6.5.

4.7.3 Spectral coherence – Correlation in the frequency domain

Just as spectral analysis (§4.5.2) reveals characteristic oscillatory behavior in a single station's climate record, cross-spectral analysis asks if 2 stations' records co-vary at similar frequencies.[125] Cross-spectra show in which frequency bands station temporal dynamics are highly correlated (spectral coherence, h^2) and if this coordination is in or out of phase (Figure 22). Coherence is the squared correlation coefficient for a given frequency band (von Storch and Zwiers 2001) and so, in a way similar to regression R^2 (§3.4.3.2), can be interpreted as the percent variance shared by the two series in that range of frequencies.

---

[122] An online facility for creating correlation maps is http://www.esrl.noaa.gov/psd/data/correlation/. While intended for hemispheric teleconnection correlations (§4.8.2), it can be adapted to creating site point-correlation maps by inserting a site station record as the custom timeseries (http://www.esrl.noaa.gov/psd/data/correlation/custom.html). See help and instruction in the page's left frame.

[123] The semivariogram (γ) for variable z is: $\gamma(h) = \frac{1}{2n(h)} \sum [z(x) - z(x+h)]^2$, where x is any location, h is the separation distance between pairs of data points, and $n(h)$ is the number of pairs that are separated by h; summation is over all such pairs. This equation expresses γ as a function of h (Figure 21c). Note that γ is calculated as *half* the spatial variance, hence '*semi*-variogram.'

[124] Note that semivariance and autocorrelation reflect the same spatial process, but in the opposite manner: as stations become less related with distance, semivariance increases (Figure 21c), while autocorrelation decreases (Figure 11b, d).

[125] Cross-spectral analysis was first introduced in the context of comparing timeseries of two variables (§4.4.1, §4.4.2).

Techniques for cross-spectral analysis include Fourier and wavelet approaches (Ghil et al. 2002, Whitcher et al. 2000).[103] Such analyses are most useful if they test for statistical significance in coherence and phase (Figure 22b, c).

4.8 Spatial patterns. II: Hemispheric teleconnections

As our ability to monitor the global climate system has grown over the last half century, we have become increasingly aware of connections between local climate variability and remote global-scale atmospheric and ocean dynamics. Such 'teleconnections' are fundamental to our understanding the roots of regional interannual climate variability and to exploring the mechanisms by which hemispheric processes scale down to the ecology of species and landscapes (cf. Stenseth et al. 2002, Graumlich et al. 2003, Stenseth and Mysterud 2005, Pederson et al. 2006).[126]

In this section, I briefly introduce well-recognized teleconnections (§4.8.1), present analytical approaches (§4.8.2), and discuss key implementation concerns (§4.8.3).

4.8.1 Modes of variation in hemispheric circulation – Multiyear oscillations

Predominant teleconnections for North America have their source in four major interannual to multidecadal oscillations of the climate system:

- El Niño-Southern Oscillation (ENSO)[127,128]
- Pacific Decadal Oscillation (PDO)[129,130]
- Northern Annular Mode (NAM)/North Atlantic Oscillation (NAO)[131]
- Atlantic Multidecadal Oscillation (AMO)[132]

These climate system oscillations have characteristic centers of action typically in places where there is strong coupling between the ocean and atmosphere (e.g., Figure 23a, *top*). These centers are quasistationary in their location, constrained by ocean basin geometry and basin-wide ocean circulation. The centers have characteristic quasiperiodic, multiyear modes of behavior in both ocean and atmospheric measures [e.g., sea surface temperatures (SST), sea level pressure (SLP);

---

[126] Comparable analyses for marine ecosystems include: Hare and Mantua (2000; with some terrestrial measures) and Schwing et al. (2009)

[127] General ENSO reference: Trenberth (1997). See also UCAR tutorial webcasts re ENSO under: http://www.nws.noaa.gov/om/csd/pds/PCU2/IC2.4.shtml (free registration, login)

[128] For looking at El Niño-related Pacific Basin dynamics but in the Northern Hemisphere, a counterpart to the Southern Oscillation (SO) is the Northern Oscillation (Schwing et al. 2002). However, indices for the two oscillations show similar dynamics and, in the literature, the SO Index (SOI; Table 4) remains the more common of the two for North American analyses. (cf. http://www.pfeg.noaa.gov/products/PFEL/modeled/indices/NOIx/noix_bkgrnd.html)

[129] General PDO reference: Mantua et al. (1997). Re the PDO and what is called the North Pacific Oscillation (NPO): The term "North Pacific Oscillation" is unfortunately used in the literature to refer to two distinct interannual dynamics in the North Pacific – either as (1) equivalent to the PDO (Gershunov and Barnett 1998) or (2) an oscillation whose spatiotemporal pattern is orthogonal (independent) of the PDO (Minobe and Mantua 1999). The first pattern predominately affects climates across North America (Trenberth and Hurrell 1994, Hurrell 1996, Mantua et al. 1997), the second pattern influences climates of the western Pacific (Linkin and Nigam 2008). Neither of these should be confused with the Northern Oscillation.[128]

[130] Much discussion re the PDO focuses on multidecadal regime shifts at roughly 15-25 and 50-70 year cycles (Minobe 1997, 1999). In addition, the PDO Index (Table 4) has substantial interannual variability (Figure 23a) – index fluctuations in and out of + or – territory can seen as oscillations within a PDO regime. However, whether regime shifts are actually characteristic of the PDO is evaluated using the paleorecord by Gedalof et al. (2002).

[131] The NAO is considered part of NAM dynamics and so are grouped together. The NAM is also referred to as the Arctic Oscillation (AO). General reference for NAO: Hurrell et al. (2003); for NAM/AO: Thompson and Wallace (2000)

[132] See Schlesinger and Ramankutty (1994), McCabe et al. (2004)

Figure 23a, b]. These oscillations are controlled by long-acting, geographically-broad interactions between atmospheric and ocean circulations and run deep vertically in both systems. These dynamics are not fully understood, however.[133] For brief overviews of these oscillation systems, their centers of action, temporal dynamics, and teleconnections, see Stenseth et al. (2003).[134]

For land and inland-waters resource science applications, we are interested in these oscillations' teleconnections to the climates of continents adjacent to (or of ocean islands embedded in) corresponding oceans, but removed from the centers of action. These teleconnections are the downstream consequences of location and strength changes in the centers-of-action's (1) warm and cool pools of ocean water and (2) semipermanent high and low pressure systems in the lower troposphere. Consequences are shifts in the position and intensity of the Intertropical Convergence Zone, Mid-latitude Jets, and Subtropical and Polar Highs – and, linked to these, of tropical and mid-latitude storm tracks, summer monsoons, winter advection of warm, moist or cold, dry air masses, and marine layer stability (e.g., Dai et al. 1998, Castro et al. 2001; Figure 24).

Table 3 provides entry links for websites displaying teleconnection patterns for surface air temperature, precipitation, and other station variables in terms of their means and extremes.

4.8.2 Testing for teleconnections – Hemispheric circulation indices

For these oscillations, circulation indices are used to represent their dynamics in a single timeseries. These portray the principal mode of variability at a center of action (e.g., SLP anomalies in the Aleutian Low pressure center, for the PDO-related North Pacific Index, NPI – Trenberth and Hurrell 1994; Table 3) or in the difference between dipoles (e.g., Tahiti-Darwin SLP difference for the Southern Oscillation Index, SOI – Figure 23b, *top*). Table 4 presents indices for oscillations with major teleconnections across North America along with download links. Take care that candidate index datasets for your analysis are not smoothed series because of problems they present in statistical evaluation (§4.5.1: Additional cautions).[135]

To explore teleconnection signals in local and regional climates, common techniques are:

Correlation approaches. These use linear correlation to evaluate the relationship between station records and a circulation index. Techniques include:

- Simple linear correlation between a station timeseries and the index (§4.4.1).
- Cross-correlation (§4.4.1) over, for example, weekly, monthly, or seasonal lags based on known or hypothesized mechanics (e.g., Barton and Ramirez 2004, Wright and Calderón 2006).[92]
- Point-correlation maps (§4.7.1), with the index as the point timeseries (Figure 24a). Teleconnection maps, as on websites in Table 3, are commonly based on this method (e.g., Castro et al. 2001). NOAA provides an on-line facility[122] for generating index-correlation maps with built-in climate fields (demonstrated in Figure 24a).

Cross-spectra/Cross-wavelet. High spectral coherence (§4.7.3) between timeseries for a station and an index may indicate that local variability is dynamically linked to the corresponding hemispheric oscillation. The phase and frequency bands of this coherence may suggest or lend

[133] In addition, these dynamics are not considered to be genuinely captured by today's global climate models, with implications re our ability to understand how these key modes of climate variability may change under future climates.
[134] See also: Steward (2005): http://oceanworld.tamu.edu/resources/oceanography-book/oceananddrought.html
[135] Also see caveats regarding smoothing in timeseries comparisons (§4.4.2) and spectral analysis (§4.5.2).

support to hypotheses for such a mechanism. Cross-wavelet analyses can also be applied with these same objectives (§4.7.3).

Data blocking.[136] Dividing an index timeseries into groups (blocks) of years can add power to statistical tests for teleconnections. This is usually done by oscillation phase – positive vs. negative phase years – or, in addition, by eliminating years with low or neutral signal (e.g., Figure 23b: *top*).[137] Blocking by phase focuses an analysis on regime state rather than intensity; blocking out near-neutral conditions focuses on strong episodes most likely to have detectable downstream effects (cf. next section, §4.8.3: *Forcing strength*).

Two-factor blocking, such as by the phases of two indices, allows for an assessment of their interaction (e.g., Figure 23b: *bottom*).[138] Alternatively, the second or yet additional factors can be any variable known or hypothesized to alter teleconnections (such as season, response region). Analysis of teleconnections with multifactor-blocked data can be by individual combinations of the blocks (e.g., the four combinations of PDO and AMO +/– phases in Figure 24b) or with all considered simultaneously, as in multifactor analysis of variance (MANOVA).[139]

Regime shift detection. A reflection of circulation regime shifts in timeseries of landscape climatic, hydrologic, and ecological variables can provide evidence of the impact of hemispheric climate processes on local dynamics (Hare and Mantua 2000; Figure 24c). Techniques for regime shift detection are referenced in §4.5.3. As with local climate data, check for temporal inconsistencies created in the processing of index timeseries you're evaluating (e.g., Table 4: footnote 154).

Principal component analysis. Generally speaking, principal component analysis (PCA) is used to describe the multivariate behavior of a system as a much reduced set of variables. For example, Hare and Mantua (2000) used PCA on an array of environmental variables to distill out principal components strongly related to PDO dynamics. In geophysics, PCA is commonly referred to as *empirical orthogonal function* (EOF) analysis, where it is applied to spatiotemporal data for one parameter (such as temperature anomalies).[140] This technique extracts the most prevailing patterns in space and their corresponding timeseries (e.g., Figure

---

[136] Blocking was introduced earlier, in the context of looking for interaction between slow vs. fast-moving processes: §4.4.2: *Process timescales differ, Mode 2*.

[137] Note that delineation of an oscillation's phases can be an issue: there is no clear or consistent definition of what magnitude or duration of change constitutes a shift to the opposite phase and there is no single index or other definitive measure of these dynamics in terms of geography, variable, or season (hence multiple indices per oscillation in Table 4, Figure 23b; see also points re PDO phases in footnote 138).

[138] An example 2-way blocking by ENSO and PDO phases is given by JISAO (University of Washington): http://www.cses.washington.edu/cig/pnwc/compensopdo.shtml. Some points regarding their presentation:

- PDO phases are presented both in terms of multidecadal regime (columns) and annual state (cell entries).[130]
- Before 1925, there is reduced certainty in climate records used to determine PDO phase. From 1900-1924, the phase is variously considered as negative (in the JISAO table; also Mantua et al. 1997) or positive (e.g., Biondi et al. 2001 based on proxy records: http://www.ncdc.noaa.gov/paleo/pubs/biondi2001/biondi2001.html; Rodionov and Overland 2005: cf. Figure 16b).
- Since 1976/77, shortterm multiyear excursions to negative and back to positive territory have been suggested as possible regime changes (e.g., 1999, 2003, 2006 shifts in the JSIAO table, 1989 shift in Hare and Mantua 2000; see also Rodionov and Overland 2005: Figure 16b).[137] The perspective of additional decades is needed to judge whether these are true regime shifts or reflect year-to-year variation within the positive phase that started in 1977 (see Figure 23a, *bottom*).

[139] See Helsel and Hirsch (2002: §7.2.2)[9]

[140] EOF analysis can be implemented using eigenvalue decomposition (eigenanalysis) or singular value decomposition (SVD); see von Storch and Zwiers (2001: §13.2.9).

23a *top* and *bottom*, respectively, for the first component of the PDO).[141] In interpreting EOF's, the first few functions (principal components) tend to explain a sufficient portion of the variance to warrant exploring their scientific meaning; the remaining EOF's are usually only of minor significance.[142]

As a key method, EOF's reveal primary modes in teleconnections (e.g., Wallace and Gutzler 1981). A bivariate technique related to EOF analysis is *canonical correlation analysis* (CCA); this approach extracts spatiotemporal patterns common to two spatially and temporally distributed parameters. von Storch and Zwiers (2001: Chapters 13, 14) present these pattern detection techniques.

4.8.3 Properties of circulation oscillations and their indices – Insights and pitfalls

Circulation indices offer us the opportunity to understand linkages from the global to local. Exploring this scale translation requires insights into the nature of circulation oscillations and their indices. These insights can help design analyses, seek mechanisms in interpretation, and avoid pitfalls.

Key features of oscillations and indices are:

Season dependence. Dynamics at centers of action fluctuate seasonally, and their downstream connections follow suit. As a result, seasonal indices tend to best capture teleconnections. Annual indices often blur the controlling signal, while individual monthly values divide seasons up arbitrarily, weakening detection of a seasonal effect.[143,144] The key teleconnective season (and how it is defined) may depend on the site climate variable evaluated, such as for temperature vs. precipitation.[145] In the extratropics, the strongest signal-to-noise ratio is commonly in winter, giving the most robust teleconnections.[146] However, it can be more fruitful to take the season that corresponds to timing of key local dynamics, such as a summer index for connections with the summer monsoon (e.g., Castro et al. 2001). On the other hand, between-season interactions might alternatively suggest evaluating an index from a different season – for example, if summer moisture conditions are more a function of winter snowpack and its melt regime, then a winter or spring index may be more appropriate than a summer one.

Forcing strength makes for a teleconnection. Strong events are most likely to propagate downstream and result in signals relevant to local dynamics. Weak signals generally get dissipated en route and swamped locally by other sources of variability. As a result,

---

[141] EOF analysis terminology (with corresponding PCA terms):
- Empirical orthogonal functions represent the spatial pattern (principal component loadings)
- EOF coefficients express the temporal pattern (principal component scores)

[142] Note that in such a spatiotemporal analysis, high percent-variance explained (in the first EOF's) ideally comes from the temporal dynamics of the entire spatial field being explained moderately well. However, note that relatively high variance explained (often in subsequent EOF's) can also come from dynamics of a restricted part of the domain being explained extremely well – this describes how one area is behaving but does little to show how different regions are connected.

[143] This is due to within-seasonal variability either (1) because a seasonal effect may be more concentrated in one month one year but in an adjacent month in the next occurrence or (2) because month boundaries are arbitrary – what's taken to be a month's difference in timing may only be a matter of a few days.

[144] Rather than using a standard scheme to define seasons (e.g., winter = December-January-February), it can be more powerful to delineate seasons based on breaks in system dynamics. For example, Trenberth and Hurrell (1994) define a winter NPI covering November through March (http://www.cgd.ucar.edu/cas/jhurrell/npindex.html).

[145] Plotting monthly teleconnection results can suggest an optimal delineation of key seasons and if that differs for different local variables (e.g., http://cses.washington.edu/cig/pnwc/clvariability.shtml#figure5).

[146] For an example from the Pacific Northwest, see http://cses.washington.edu/cig/pnwc/clvariability.shtml#figure1.

teleconnections are most evident when analyses contrast the strongest periods of opposite phases (e.g., by blocking out near-neutral periods, discussed in §4.8.2: *Data blocking*) (Figure 23b, *top*).

Basin interactions. Circulation oscillations arising in different ocean basins or in different sectors of a basin interact, such as between the PDO and AMO (McCabe et al. 2004) and ENSO and PDO (Gershunov and Barnett 1998), respectively. This leads to conditional teleconnections, where the phase of one oscillation influences the downstream expression of another (e.g., Figure 24b). Two-factor blocking and related multifactor techniques are approaches for evaluating interactions (cf. §4.8.2: *Data blocking*).

Circulation indices are broadly integrative. Oscillation indices track climate dynamics of broad regions of the globe that have strong spatiotemporal coherence across many variables (Figure 23a). Stenseth and Mysterud (2005) layout a conceptual framework for how these hemispheric, seasonal, and multivariate indices present an integrated view of climate that sets up local conditions for a season or longer. These establish prevailing seasonal conditions in ways that can have as much power in explaining ecological dynamics as do analyses of local weather (e.g., Hallett et al. 2004, Forchhammer and Post 2004).[147]

Additional considerations in scale linkages to landscapes and species are:

Local conditionality. How forcing from a given teleconnection plays out across landscapes and regions can be conditional on physiographic features such as aspect, altitude, and latitude. This is especially the case if teleconnections affect where a critical weather threshold, such as a storm's snowline, crosses the domain (Stenseth and Mysterud 2005).

Indirect ecological effects. In evaluating hemispheric linkages to ecological dynamics (while skipping over local climate), keep in mind that some consequences of circulation teleconnections may be indirect, possibly with strong temporal lags and spatial offsets. These may arise from population and trophic dynamics and from biogeographic linkages (e.g., for regional or hemispheric migrants[148]) (Forchhammer and Post 2004).

Important caveats and common pitfalls in teleconnection analyses include:[149]

Responses nonlinear. Within a given oscillation's phase, teleconnections are not expressed the same way in each occurrence. The relationship between an oscillation's phase and its local teleconnections may in fact change sign between lower forcings vs. higher ones. This is because oscillation dynamics often shift the position of an atmospheric circulation system (e.g., storm track latitudes) into a region at first and, subsequently under a more intense teleconnection, push the circulation farther along but out of the region. Such non-monotonic, complex dynamics are of course not adequately explored with linear correlation methods. Careful period blocking can help reveal these dynamics.

Responses nonstationary. Teleconnections also appear to change with time. Some nonstationarity can be attributed to behavior conditional on the phase of other oscillation

[147] *N.B.* – Descriptions of circulation patterns in Forchhammer and Post (2004) are not entirely accurate; Stenseth et al. (2003) provide a better review for ecological audiences. Forchhammer and Post (2004) do, however, present key insights from three case studies on teleconnections and ecological dynamics.
[148] For example, when migrants have distant seasonal ranges strongly affected by a teleconnection seen there, but not in the local study domain.
[149] Stenseth et al. (2003) also review benefits and drawbacks of teleconnection analyses.

systems, as noted earlier. Aside from such conditionality, keep in mind that key teleconnections found may not have held throughout the historical period nor persist in the future (e.g., Gedalof et al. 2002).

Prediction. Much attention has been given to the prospect of predicting local climate and ecological dynamics based on teleconnections. While teleconnections with local dynamics may be statistically significant, the percent variance explained may put into question the utility of relationships for prediction.[150] Low explained variance comes from various sources:

- Nonlinearity, nonstationarity, and indirect effects of teleconnections that reduce their detection.
- Centers of action for oscillations are much removed and their signal is altered in transit by unaccounted-for downstream climate processes – including stochastic and chaotic behavior inherent to the climate system.
- Hemispheric circulation dynamics are only part of the local story – Other, independent local factors also control site climate and ecological behavior.

4.9 Interpretation of results – The good, the bad, and the ugly

4.9.1 Statistical interpretation

On completing analyses, comes interpretation of descriptive and statistical results and their scientific review (next section, §4.9.2) (Figure 1c). Recall your research questions and corresponding stated hypotheses – in doing so, confirm that your analyses are addressing what you intended. For statistical tests, review results and their significance level, and formally state outcomes (e.g., rejecting null hypotheses or not).

4.9.2 Scientific interpretation of statistical results

Caution is needed in interpreting both significant and nonsignificant statistical results:

Significant results. Statistically-significant results lend support to your hypothesized dynamics, but not to validation. In tests comparing variables, keep in mind, as noted earlier, the proposition that "correlation does not imply causation" (§4.4.2: *Interpretation of correlation*). There may be other underlying mechanisms that give rise to the relationships you see. In such bivariate as well as spatial and temporal relationships, significant patterns may turn out be nonstationary – that is, only present under conditions set by some undetected overriding process (e.g., a seemingly persistent circulation regime). Given a longer timeframe, relationships seen today may shift or disappear.[151] The role of undetected, unevaluated factors can interfere with seeking mechanisms in the interpretation of results.

Weak results. Nonsignificant results also need to be evaluated with care. Weak results (tendencies consistent with a hypothesis, but not backed by statistical significance) should not be reported as apparent support for your hypothesis. Rather they support re-evaluation of:

- Available data – Are additional, appropriate data sources available that would lengthen a station's record, or allow evaluation across several locations? The advantage of this is that the power of statistical tests increases with number of independent observations.

[150] The distinction here is between significance level (e.g., $p<0.05$) and % variance explained, e.g., R^2 for regression analysis (§3.4.3.2). It is not unusual for teleconnection analysis to yield highly significant regressions, say $p<0.01$, but with it low to modest % variance explained, commonly with R^2's<0.30.
[151] See Schumm (1991: Chapter 3) for pitfalls in interpreting results in space and time.

- Data processing – Review if dataset development methods adequately corrected problems or, instead, obscured patterns being testing for.
- Analysis methodology – Some techniques are more powerful than others in detecting patterns. Review analysis options (data transformations, alternate tests, etc.) and check that their assumptions are followed.

Clearly nonsignificant results. From a statistical perspective, clearly nonsignificant results obviously do not support your hypothesis. From a scientific perspective, this can be the end of story or lead to alternate hypotheses. However, another proposition that "correlation is needed to prove causation" is not necessarily the case. We may just not be looking at the question from the right perspective or with the right tools: system complexity can mask causation, making it recalcitrant to standard approaches. Climate processes and affected biotic components (1) are highly interactive with positive and negative feedbacks, (2) act across a range of time and space scales, and (3) often entail non-monotonic or threshold responses. Numerical simulation models are tools employed to understand dynamics of such highly connected systems.

5.0 Synopsis

The objectives of this report are two-fold. First, to layout a methodology for developing quality climate datasets appropriate for resource management science. Second, to introduce the array of analysis techniques available for answering the many questions we often ask regarding climate dynamics and their interaction with a region's ecology and other landscape processes.

Overarching guidelines for creating datasets are:

(1) *Hypotheses dictate the requirements* of datasets, selection of data clean-up methodologies, and corresponding analyses. Keeping hypotheses in mind throughout the process will help ensure a successful outcome: a credible dataset and valid results.

(2) *No dataset is perfect.* There's a trade off between working with a clean highly-processed data set and one that is as unadulterated as possible.

   (a) *No raw observational dataset is free of collection and archive errors*. Decisions on what types of errors to be concerned about and to correct should be based on goals and analysis requirements. Spending effort on ridding a dataset of all problems may not be necessarily for intended uses, and may create a set that is not appropriate for addressing some questions. Keep in mind that some adjustment techniques may interfere with your intended analysis.

   (b) *No cleaned-up dataset is free of assumptions* about what its planned or perceived use is. Keep track of decisions made along the way and limitations arising as a consequence of these decisions and processing. Test generated data for unintended emergent features, and that these aren't going to create spurious results in the analysis stage.

(3) *Nonetheless, techniques for improving dataset utility can detect and correct* data errors, biases, and artificial record inhomogeneities and infill missing observations.

(4) *Document achieved improvements, intended uses, and caveats* for yourself and other users to understand what the dataset is good vs. inappropriate for.

Strategies for data analysis are:

(5) *Analyses must include statistical significance testing* to have results that can be relied upon for understanding your system and for decision-making. This is in addition to analyses being well designed and properly interpreted. Descriptive and graphic techniques are valuable for exploring data, but ultimately these must lead to hypothesis generation and testing.

(6) *Statistical techniques make certain assumptions about input data.* For a given analysis, some methods have more relaxed requirements than others regarding statistical distribution, independence, and missing values, for example. Keep track of the properties of your data with respect to analysis requirements – violating these will most likely give overestimated significance to your results.

(7) *Analytical tools can evaluate temporal and spatial patterns at different scales,* including trends, oscillations, regime shifts, daily events, regional correlations, and hemispheric teleconnections.

(8) *Keep in mind that results contain uncertainties not assessed statistically.* Uncertainties external to an analysis will mean we're likely to assign too much confidence to statistical results. Sources of uncertainty include unresolved data issues, analysis limitations, and simplifying assumptions re natural systems that we use to guide our studies.

Lastly:

(9) *Studies are most likely to be successful if driven by a sense for underlying mechanisms* that connect ecosystems and species with climate. This will provide more confidence in interpretation of analysis results beyond just their statistical basis, adding insights into the dynamics of park landscapes.

Acknowledgements

Many thanks to John Gross, Stephen Gray, Mark Williams, Brent Frakes, Stacey Ostermann-Kelm, Cathie Jean, Chris Daly, Chris Knud-Hansen, Mark Otto, and anonymous reviewers for suggestions and comments. Thanks also to Michael Hartman and Todd Ackerman who contributed to Niwot Ridge climate analyses presented in some of the figures. This report was funded by the National Park Service, Rocky Mountains Cooperative Ecosystem Studies Unit (RM-CESU) in support of the Inventory and Monitoring Program. This publication is in addition a contribution of the Niwot Ridge Long-Term Ecological Research (LTER) project funded by the National Science Foundation.

References

Alexander, L.V., et al. [23 coauthors]. 2006. Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res*. 111, D05109, doi:10.1029/2005JD006290.

Allen, T.F.H., and T.B. Starr. 1982. *Hierarchy: Perspectives for Ecological Complexity.* Chicago: The University of Chicago Press. 310 p.

Alley, W.M. 1984. The Palmer Drought Severity Index: limitations and assumptions. *Journal of Climate and Applied Meteorology* 23:1100-1109

Banerjee, S., B.P. Carlin, and A.E. Gelfand. 2003. *Hierarchical Modeling and Analysis for Spatial Data.* Monographs on Statistics and Applied Probability, Routledge. 472 p.

Baron, J.S. 2006. Hindcasting nitrogen deposition to determine an ecological critical load. *Ecological Applications* 16:433-439

Barton, S.B., and J.A. Ramirez. 2004. Effects of El Niño Southern Oscillation and Pacific Interdecadal Oscillation on water supply in the Columbia River Basin. *J. Water Resources Planning and Management* 130:281-289. doi:10.1061/(ASCE)0733-9496(2004)130:4(281)

Beniston, M., and D. Stephenson. 2004. Extreme climatic events and their evolution under changing climatic conditions. *Global and Planetary Change* 44:1-9

Billings, W.D. and L.C. Bliss. 1959. An alpine snowbank environment and its effects on vegetation, plant development, and productivity. *Ecology* 40:388-397

Biondi, F., A. Gershunov, and D.R. Cayan. 2001. North Pacific decadal climate variability since 1661. *J. Climate* 14:5–10

Biondi, F., T.J. Kozubowski, and A.K. Panorska. 2002. Stochastic modeling of regime shifts. *Climate Research* 23:23–30

Box, G.E.P., and G.C. Tiao. 1975. Intervention analysis with applications to economic and environmental problems. *J. Amer. Stat. Assoc.* 70:70–79

Braswell, B.H., D.S. Schimel, E. Linder, B. Moore III. 1997. The Response of Global Terrestrial Ecosystems to Interannual Temperature Variability. Science 278:870-873

Burnaby, T.P. 1953. A suggested alternative to the correlation coefficient for testing the significance of agreement between pairs of time series, and its application to geologic data. *Nature* 172:210–211

Burroughs, J. 2008. *Integrated Global Radiosonde Archive - Quality Control.* National Climate Data Center, NOAA. Web document: http://www.ncdc.noaa.gov/oa/climate/igra/index.php?name=quality

Cade, B.S., and B.R. Noon. 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* 1:412-420

Castro, C.L., T.B. McKee, and R.A. Pielke, Sr. 2001. The relationship of the North American monsoon to tropical and North Pacific sea surface temperatures as revealed by observational analyses. *J. Climate* 14:4449-4473

Clark, I., and W.V. Harper. 2000. *Practical Geostatistics 2000.* Ecosse North America. 442 p.

Chang, Y.S., D. Jeon, H. Lee, H.S. An, J.W. Seo, and Y.H. Youn. 2004. Interannual variability and lagged correlation during strong El Niño events in the Pacific Ocean. *Climate Research* 27:51-58

Chatfield, C. 1995. *Problem Solving: A Statistician's Guide*, 2nd ed. Chapman & Hall. 350 p. Available in part on Google Books: http://books.google.com/books?id=EA3jBSe0c3wC

Chen, M., W. Shi, P. Xie, V.B.S. Silva, V.E. Kousky, R.W. Higgins, and J.E. Janowiak. 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res*. 113, D04110, doi:10.1029/2007JD009132. 13 p.

Christy, J.R., W.B. Norris, and R.T. McNider. 2009. Surface temperature variations in East Africa and possible causes. *J. Climate* 22**:**3342–3356

CLIMAS. 2002. Climate Divisions: to use or not to use? *END InSight,* November 2002, p. 3-4. CLIMAS (Climate Assessment for the Southwest), University of Arizona. http://www.climas.arizona.edu/forecasts/articles/climdiv_nov2002.pdf

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values.* Springer. 225 p.

Conover, W.J. 1999. *Practical Nonparametric Statistics*, 3rd ed. Wiley. 584 p.

Crawley, M.J. 2002. *Statistical Computing: An Introduction to Data Analysis using S-Plus.* Wiley. 772 p.

Crawley, M.J. 2007. *The R Book.* Wiley. 950 p.

Cressie, N.A.C. 1993a. *Statistics for Spatial Data.* Wiley, New York. 928 p.

Cressie. N. 1993b. Aggregation in geostatistical problems. In: Soares, A. (ed) *Geostatistics Troia '92*, Vol 1, pp. 25–36. Kluwer Academic Publishers, Dordrecht.

Dai, A., K.E. Trenberth, and T.R. Karl 1998. Global variations in droughts and wet spells: 1900-1995. *Geophysical Research Letters* 25(17):3367-3370

Daly, C., W.P. Gibson, G.H. Taylor, G.L. Johnson, and P. Pasteris. 2002. A knowledge-based approach to the statistical mapping of climate. *Climate Research* 22:99-113 http://www.prism.oregonstate.edu/

Daly, C., W.P. Gibson, G.H. Taylor, M.K. Doggett, and J.I. Smith. 2007. Observer bias in daily precipitation measurements at United States Cooperative Network Stations. *Bull. Amer. Meteor. Soc.* 88**:**899–912.

Daly, C., M. Halbleib, J.I. Smith, W. P. Gibson, M. K. Doggett, G.H. Taylor, J. Curtis, and P.P. Pasteris. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology* 28:2031-2064 http://www.prism.oregonstate.edu/

Delcourt, H.R., P.A. Delcourt, and T. Webb III. 1983. Dynamic plant ecology: the spectrum of vegetation change in space and time. *Quat. Sci. Rev.* 1:153-175

Díaz, S.C., C.A. Salinas-Zavala, S. Hernandez-Vazquez. 2008. Variability of rainfall from tropical cyclones in northwestern Mexico and its relation to SOI and PDO. *Atmósfera* 21: 213 - 223

Di Luzio, M., G.L. Johnson, C. Daly, J.K. Eischeid, and J.G. Arnold. 2008. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *J. Appl. Meteor. Climatol.* 47:475–497

Doesken, N.J. 2005. The National Weather Service MMTS (Maximum-Minimum Temperature System) -- 20 years after. *15th Conference on Applied Climatology and 13th Symposium on Meteorological Observations and Instrumentation.* Paper JP1.26. America Meteorological Society, Boston. http://ams.confex.com/ams/pdfpapers/91613.pdf

Draper, N.R., and H. Smith. 1998. *Applied Linear Regression,* 3rd ed. Wiley. 736 p.

Drews, C. 2003. *Detecting Climate Change in Canadian Ice Data.* Research Paper, USDA Graduate School. http://www.highestlake.com/canadice.html

Easterling, D.R., T.R. Karl, E.H. Mason, P.Y. Hughes, D.P. Bowman, R.C. Daniels, and T.A. Boden. 1996. *United States Historical Climatology Network (US HCN) Monthly Temperature and Precipitation Data.* ORNL/CDIAC-87, NDP-019/R3. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory. http://www.ncdc.noaa.gov/oa/climate/research/ushcn/ushcn.html. Data access: http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html

Easterling, D.R., B. Horton, P.D. Jones, T.C. Peterson, T.R. Karl, D.E. Parker, M.J. Salinger, V. Razuvayev, N. Plummer, P. Jamason, and C.K. Folland. 1997. Maximum and minimum temperature trends for the globe. *Science* 277:364-367

Easterling, D.R., T.R. Karl, J.H. Lawrimore, and S.A. Del Greco. 1999. *United States Historical Climatology Network Daily Temperature, Precipitation, and Snow Data for 1871-1997.* ORNL/CDIAC-118, NDP-070. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee. http://www.ncdc.noaa.gov/oa/climate/research/ushcn/ushcn.html. Data access: http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html

Easterling, D.R., J.L. Evans, P.Ya. Groisman, T.R. Karl, K.E. Kunkel, and P. Ambenje. 2000. Observed variability and trends in extreme climate events: a brief review. *Bulletin of the American Meteorological Society* 81:417–425

Eischeid, J.K., C.B. Baker, T.R. Karl, and H.F. Diaz. 1995. The quality control of long-term climatological data using objective data analysis. *J. Appl. Meteor.* 34:2787–2795

Endoe, J. 2009. *The USHCN Version 2 Serial Monthly Dataset.* National Climate Data Center, NOAA. http://www.ncdc.noaa.gov/oa/climate/research/ushcn/

Forchhammer, M.C., and E. Post. 2004. Using large-scale climate indices in climate change ecology studies. *Population Ecology* 46:1-12

Frei, C., and C. Schär. 2001. Detection probability of trends in rare events: Theory and application to heavy precipitation in the Alpine region. *Journal of Climate* 14:1568-1584

Fuentes, M., T.G.F. Kittel, and D. Nychka. 2006. Sensitivity of ecological models to their climate drivers: Statistical ensembles for forcing. *Ecological Applications* 16:99–116

Gandin, L.S. 1988. Complex quality control of meteorological observations. *Monthly Weather Review* 116:1137-1156

Garson, G.D. 2009. *Statnotes: Topics in Multivariate Analysis.* North Carolina State University http://www2.chass.ncsu.edu/garson/pa765/statnote.htm

Gedalof, Z., and D.J. Smith. 2001. Interdecadal climate variability and regime-scale shifts in Pacific North America. *Geophys. Res. Lett.* 28:1515–1518

Gedalof, Z., N.J. Mantua, and D.L. Peterson. 2002. A multi-century perspective of variability in the Pacific Decadal Oscillation: new insights from tree rings and coral. *Geophys. Res. Lett*. 29(24): 2204, doi:10.1029/2002GL015824.

Gershunov, A., and T. Barnett. 1998. Interdecadal modulation of ENSO teleconnections. *Bulletin of the American Meteorological Society* 79(12):2715-2725.

Ghil, M., M.R. Allen, M.D. Dettinger, K. Ide, D. Kondrashov, M.E. Mann, A.W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou. 2002. Advanced spectral methods for climatic time series. *Rev. Geophys.* 40(1), 1003, doi:10.1029/2000RG000092

Gleason, K.L., J.H. Lawrimore, D.H. Levinson, T.R. Karl, and D.J. Karoly. 2008. A revised U.S. Climate Extremes Index. *J. Climate* 21:2124-2137 http://www.ncdc.noaa.gov/oa/climate/research/cei/cei.html

Goubanova, K., and L. Lia. 2007. Extremes in temperature and precipitation around the Mediterranean basin in an ensemble of future climate scenario simulations. *Global and Planetary Change* 57:27-42

Graumlich L.J., M.F.J. Pisaric, L.A. Waggoner, J.S. Littell, and J.C. King. 2003. Upper Yellowstone River flow and teleconnections with Pacific Basin climate variability during the past three centuries. *Climatic Change* 59:245-262

Gray, S. T., J. L. Betancourt, C. L. Fastie, and S. T. Jackson. 2003. Patterns and sources of multidecadal oscillations in drought-sensitive tree-ring records from the central and southern Rocky Mountains. *Geophys. Res. Lett.* 30(6):1316, doi:10.1029/2002GL016154. 4 p.

Guttman, N. B. 1999. Accepting the Standardized Precipitation Index: a calculation algorithm. *J. Amer. Water Resour. Assoc.* 35(2):311-322

Guttman, N., and R. Quayle. 1996. A historical perspective of U.S. climate divisions. *Bulletin of the American Meteorological Society* 77(2):293–304

Haas, T.C. 1995. Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *J Am Stat Assoc* 90:1189-1199

Haimberger, L. 2007. Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate* 20:1377–1403

Hale, R.C., K.P. Gallo, T.W. Owen, and T.R. Loveland. 2006. Land use/land cover change effects on temperature trends at U.S. Climate Normals stations. *Geophys. Res. Lett.* 33, L11703, doi:10.1029/2006GL026358

Hallett, T.B., T. Coulson, J.G. Pilkington, T.H. Clutton-Brock, J.M. Pemberton, and B. Grenfell. 2004. Why large-scale climate indices seem to predict ecological processes better than local weather. *Nature* 430:71–75

Hamed, K.H., and A.R. Rao. 1998. A modified Mann-Kendall test for autocorrelated data. *J. Hydrology* 204:182-196

Hansen, J., R. Ruedy, J. Glascoe, and M. Sato. 1999. GISS analysis of surface temperature change. *J. Geophys. Res.* 104(D24):30997-31022

Hao, L., and D.Q. Naiman. 2007. *Quantile Regression.* Quantitative Applications in the Social Sciences, v. 149. Sage Publications. 136 pp. In part on Google Books: http://books.google.com/books?id=nKVBXePWtSoC

Harcum, J.B., J.C. Loftis, and R.C. Ward. 1992. Selecting trend tests for water quality series with serial correlation and missing values. *Water Resources Bulletin* 28:469-478

Hare, S.R. and N.J. Mantua. 2000. Empirical evidence for North Pacific regime shifts in 1977 and 1989. *Prog. Oceanogr.* 47(2-4):103-146

Heim, R.R. 2002. A review of twentieth-century drought indices used in the United States. *Bull. Am. Meteorol. Soc.* 83:1149−1165

Helsel, D.R., and R.M. Hirsch. 2002. *Statistical Methods in Water Resources.* U.S. Geological Survey, Techniques of Water-Resources Investigations Book 4, Chapter A3. http://pubs.usgs.gov/twri/twri4a3/index.html

Herring, D. 2007. *Earth's Temperature Tracker.* Earth Observatory, NASA. http://earthobservatory.nasa.gov/Study/GISSTemperature/giss_temperature3.html

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25:1965–1978. doi: 10.1002/joc.1276

Hirsch, R.M., and J.R. Slack. 1984. A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research* 20(6):727–732

Hirsch, R.M., J.R. Slack, and R.A. Smith. 1982. Techniques of trend analysis for monthly water quality data. *Water Resources Research* 18:107-121

Holling, C.S. 1992. Cross-scale morphology, geometry and dynamics of ecosystems. *Ecol. Monogr.* 62(4):447-502

Hosking, J.R.M., and Wallis, J.R. 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge, UK. 224 p. Available in part on Google Books: http://books.google.com/books?id=OfB57wi7VcUC

Hu, Q., and G.D. Willson. 2000. Effects of temperature anomalies on the Palmer Drought Severity Index in the central United States. *Int. J. Climatol.* 20, 1899-1911

Hubbard, K.G., N.B. Guttman, J.S. You, and Z. Chen. 2007. An improved QC process for temperature in the daily cooperative weather observations. *Journal of Atmospheric and Oceanic Technology* 24:206-213

Hurrell, J.W. 1996. Influence of variations in extratropical wintertime teleconnections on Northern Hemisphere temperature. *Geophys. Res. Lett.* 23:665–668.

Hurrell, J.W., Y. Kushnir, G. Ottersen and M. Visbeck. 2003. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact.* Geophysical Monograph Series. American Geophysical Union, Washington, DC.

Hutchinson, M.F. 1995. Interpolating mean rainfall using thin plate smoothing splines. *International Journal of Geographical Information Systems* 9:385–403.

Hutchinson, M.F. 2004. *Anusplin Version 4.3.* Centre for Resource and Environmental Studies. The Australian National University: Canberra, Australia.

Imbrie, J., E. Boyle, S. Clemens, A. Duffy, W. Howard, G. Kukla, J . Kutzbach, D. Martinson, A. Mcintyre, A. Mix, B. Molfino, J. Morley, L. Peterson, N. Pisias, W. Prell, M. Raytoo, N. Shackleton, and J. Toggweiler. 1992. On the structure and origin of major glaciation cycles. 1. Linear responses to Milankovitch forcing. *Paleoceanography* 7:701-738

Imbrie, J., A. Berger, E. Bogle, S. Clemens, A. Duffy, W. Howard, G. Kukla, J. Kutzbach, D. Martinson, A. McIntyre, A. Mir, B. Molfino, J. Morley, L. Peterson, N. Pisias, W. Prell, M. Raymo, N. Shackleton, and J. Toggweiler. 1993. On the structure and origin of major glaciation cycles. 2. The 100,000 year cycle. *Paleoceanography* 8:699-735

Karl, T.R., and R.W. Knight. 1985. *Atlas of Monthly Palmer Hydrological Drought Indices (1931–1983) for the Contiguous United States*. Historical Climatology Series 3-7, National Climatic Data Center, Asheville, North Carolina, USA.

Karl, T.R., and C.N. Williams, Jr. 1987. An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Climate Appl. Meteor.* 26:1744-1763

Karl, T.R., C.N. Williams, Jr., P.J. Young, and W.M. Wendland. 1986. A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperature for the United States. *J. Climate Appl. Meteor.* 25:145-160

Karl, T.R., H.F. Diaz, and G. Kukla. 1988. Urbanization: its detection and effect in the United States climate record. *J. Climate* 1:1099-1123

Katz, R.W., and B.G. Brown. 1992. Extreme events in a changing climate: Variability is more important than averages. *Climatic Change* 21:289–302

Kestin T., D.J. Karoly, J.L. Yang, and N.A. Rayner. 1998. Time-frequency variability of ENSO and stochastic simulations. *J. Climate* 11:2258-2272

Kharin, V.V., and F.W. Zwiers. 2000. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. *J. Climate* 13:3760–3788

Kimball, S.L., B.D. Bennett, and F.B. Salisbury. 1973. The growth and development of montane species at near-freezing temperatures. *Ecology* 54:168-173

Kirkman, T.W. 1996. *Statistics to Use.* College of Saint Benedict and Saint John's University. http://www.physics.csbsju.edu/stats/

Kittel, T.G.F. 1990. Climatic variability in the shortgrass steppe. Pp. 67-75, in: D. Greenland and L.W. Swift, Jr. (eds.). *Climate Variability and Ecosystem Response*. U.S. Forest Service, Southeastern Region. Gen. Tech. Rpt. SE-65

Kittel, T.G.F., N.A. Rosenbloom, J.A. Royle, C. Daly, W.P. Gibson, H.H. Fisher, P. Thornton, D. Yates, S. Aulenbach, C. Kaufman, R. McKeown, D. Bachelet, D.S. Schimel, and VEMAP2 Participants. 2004. The VEMAP Phase 2 bioclimatic database. I: A gridded historical (20th century) climate dataset for modeling ecosystem dynamics across the conterminous United States. *Climate Research* 27:151-170

Kittel, T.G.F., T. Ackerman, M. Hartman, M.W. Williams, M.V. Losleben, and K. Chowanski. Elevational and seasonal dependence in recent climate change across an alpine and subalpine landscape in the Colorado Front Range, USA. In preparation.

Koenker, R. 2005. *Quantile Regression*. Cambridge University Press. 368 pp. In part on Google Books: http://books.google.com/books?id=Xi_dTAeAmGcC
Koenker, R., and F. Schorfheide. 1994. Quantile spline models for global temperature change. *Climatic Change* 28: 395-404

Kohler, M.A. 1949. Double-mass analysis for testing the consistency of records and for making adjustments. *Bull. Amer. Meteor. Soc.* 30:188–189

Kondrashov, D., and M. Ghil. 2006. Spatio-temporal filling of missing points in geophysical data sets. *Nonlin. Processes Geophys.* 13:151-159

Kondrashov, D., and M. Ghil. 2007. Reply to T. Schneider's comment on "Spatio-temporal filling of missing points in geophysical data sets." *Nonlin. Processes Geophys.* 14:3-4

Kulkarni, A., and H. von Stroch. 1995. Monte Carlo experiments on the effect of serial correlation on the Mann-Kendall test of trend. *Meteorol. Z.* 4(2):82-85

Kvålseth, T.O. 1985. Cautionary note about R^2. *The American Statistician* 39(4):279-285

Labat, D. 2005. Recent advances in wavelet analyses: Part 1. A review of concepts. *Journal of Hydrology* 314:275-288

Lane, D.L. 2007. *HyperStat Online Statistics Textbook.* http://davidmlane.com/hyperstat/index.html

Leffler, R., and K. Redmond. 2004. *PCU6 – Unit No. 2 Factors Affecting the Accuracy and Continuity of Climate Observations.* National Weather Service, NOAA. 40 p. http://www.weather.gov/om/csd/pds/PCU6/IC6_2/tutorial1/PCU6-Unit2.pdf (see corresponding tutorial at: http://www.weather.gov/om/csd/pds/PCU6/IC6_2/tutorial1/Factors.htm)

Legates, D.R. 1995. Global and terrestrial precipitation: A comparative assessment of existing climatologies. *International Journal of Climatology* 15:237-258

Lettenmaier, D.P., E.F. Wood, and J.R. Wallis. 1994. Hydro-climatological trends in the continental United States, 1948-88. *J. Clim.* 7:586-607

Linkin, M.E., and S. Nigam. 2008. The North Pacific Oscillation–West Pacific Teleconnection Pattern: mature-phase structure and winter impacts. *Journal of Climate* 21(9):1979-1997

Makkonen, L. 2008. Problems in the extreme value analysis. *Structural Safety* 30:405–419

Malmgren, B.A., A. Winter, and D. Chen. 1998. El Niño–Southern Oscillation and North Atlantic Oscillation control of climate in Puerto Rico. *J. Climate* 11:2713–2717

Mann, M.E. 2004. On smoothing potentially non-stationary climate time series. *Geophys. Res. Lett*. 31, L07214, doi:10.1029/2004GL019569.

Mantua, N.J., S.R. Hare, Y. Zhang, J.M. Wallace, and R.C. Francis. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of American Meteorological Society* 76:1069–1079

Marchand, P.J. 1996. *Life in the Cold.* 3rd ed. University Press of New England, Hanover, NH. 304 p.

Marlon, J.R., P.J. Bartlein, M.K. Walsh, S.P. Harrison, K.J. Brown, M.E. Edwards, P.E. Higuera, M.J. Power, R.S. Anderson, C. Briles, A. Brunelle, C. Carcaillet, M. Daniels, F.S. Hu, M. Lavoie, C. Long, T. Minckley, P. J.H. Richard, A.C. Scott, D.S. Shafer, W. Tinner, C.E. Umbanhowar, Jr, and C. Whitlock. 2009. Wildfire responses to abrupt climate change in North America. *Proc. National Academy of Science* 106: 2519-2524. Supporting information: www.pnas.org/cgi/content/full/0808212106/DCSupplemental

Martinez-Yrizar, A., and J. Sarukhan. 1990. Litterfall patterns in a tropical deciduous forest in Mexico over a five-year period. *Journal of Tropical Ecology* 6(4):433-444

Matsuo, T. 2005. Nonstationary covariance modeling using wavelets. Institute for Mathematics Applied to Geosciences (IMAGe), National Center for Atmospheric Research, Boulder CO. http://www.image.ucar.edu/GSP/Projects/ResearchNuggets.shtml#WaveletCovariance.

McCabe, G.J., M.A. Palecki, and J.L. Betancourt. 2004. Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *Proceedings of the National Academy of Sciences* 101(12):4136-4141

McDonald, J.H. 2009. *Handbook of Biological Statistics.* 2nd ed. Sparky House Publishing, Baltimore, Maryland. http://udel.edu/~mcdonald/statintro.html and http://www.lulu.com/content/3862228

McKee, T.B., N. J. Doesken, and J. Kliest. 1993. The relationship of drought frequency and duration to time scales. In: *Proceedings of the 8th Conference of Applied Climatology, 17-22 January, Anaheim, CA*. American Meteorological Society, Boston, MA. p. 179-184

Mearns, L.O., R.W. Katz, and S.H. Schneider. 1984. Extreme high temperature events: Changes in their probabilities with changes in mean temperature. *J. Climate Appl. Meteor.* 23:1601–1613.

Meehl, G.A., et al. [16 coauthors]. 2000. An introduction to trends in extreme weather and climate events: Observations, socioeconomic impacts, terrestrial ecological impacts, and model projections. *Bull. Amer. Meteor. Soc.* 81:413–416

Meek, D.W., and J.L. Hatfield. 1994. Data quality checking for single station meteorological databases. *Agricultural and Forest Meteorology* 69:85-109

Menne, M.J., and C.N. Williams, Jr. 2005. Detection of undocumented change points using multiple test statistics and composite reference series. *J. Climate* 18:4271-4286

Minobe, S. 1997. A 50-70 year climatic oscillation over the North Pacific and North America. *Geophysical Research Letters* 24:683-686

Minobe, S. 1999. Resonance in bidecadal and pentadecadal climate oscillations over the North Pacific: Role in climatic regime shifts. *Geophys. Res. Lett.* 26:855-858

Minobe, S., and N. Mantua. 1999. Interdecadal modulation of interannual atmospheric and oceanic variability over the North Pacific. *Progress in Oceanography* 43:163–192

Miranda, P.M.A., and A.R. Tomé. 2009. Spatial structure of the evolution of surface temperature (1951–2004). *Climatic Change* 93:269-284. doi 10.1007/s10584-008-9540-8

Muggeo, V.M.R. 2009. *Segmented: Segmented Relationships in Regression Models*. R Package. http://cran.es.r-project.org/web/packages/segmented/segmented.pdf

PACN. 2008. *Climate Monitoring Protocol for the Pacific Island Network (PACN), Standard Operating Procedure (SOP) # 24: Data Analysis - Instructions for Graphs and Calculations.* Version 1.0 (June 13, 2008). Pacific Island Network, National Park Service, USDI. 11 p. Unpublished document.

Palmer, W.C. 1965. *Meteorological Drought.* Research Paper No. 45, US Department of Commerce Weather Bureau, Washington, DC.

Parlange, M.B., and R.W. Katz. 2000. An extended version of the Richardson model for simulating daily weather variables. *Journal of Applied Meteorology* 39:610-622

Pederson, G., S. Gray, D. Fagre, and L. Graumlich. 2006. Long-duration drought variability and impacts on ecosystem services: a case study from Glacier National Park, Montana. *Earth Interactions* 10(1):1-28

Peterson, E.W., and L. Hasse. 1987. Did the Beaufort scale or the wind climate change. *J. Physical Oceanogr*. 17:1071-1074

Peterson, T.C., and D.R. Easterling. 1994. Creation of homogeneous composite climatological reference series. *Int. J. Climatol*. 14:671–679

Peterson, T.C., et al. 1998a. Homogeneity adjustments of *in situ* atmospheric climate data: A review. *Int. J. Climatol.* 18:1493-1517

Peterson, T.C., R. Vose, R. Schmoyer, and V. Razuvaëv. 1998b. Global historical climatology network (GHCN) quality control of monthly temperature data. *Int. J. Climatol.* 18:1169-1179

Quayle, R.G., D.R. Easterling, T.R. Karl, and P.Y. Hughes. 1991. Effects of recent thermometer changes in the cooperative station network. *Bull. Am. Meteorol. Soc.* 72:1718-1724

Redmond, K., G. McCurdy, and G. Kelly. 2008. *NPS Climate Data and Monitoring Options.* Report to the National Park Service. NPS Evaluation Report 20080518 / WRCC Report 08-01. Western Regional Climate Center, Reno, NV. 45 p.

Rehfeldt, G.E., N.L. Crookston, M.V. Warwell, and J.S. Evans. 2006. Empirical analyses of plant-climate relationships for the western United States. *International Journal of Plant Sciences* 167:1123-1150

Richardson, C.W. 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resources Research* 17:182-190

Rodionov, S.N., and J.E. Overland. 2005. Application of a sequential regime shift detection method to the Bering Sea ecosystem. *ICES J. Mar. Sci*. 62:328-332

Running S.W., R.R. Nemani, and R.D. Hungerford. 1987. Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis. *Can. J. For. Res.* 17:472-483

Schlesinger, M.E., and N. Ramankutty. 1994. An oscillation in the global climate system of period 65-70 years. *Nature* 367:723–726

Schreuder, H.T., R. Ernst, and H. Ramirez-Maldonado. 2004. *Statistical Techniques for Sampling and Monitoring Natural Resources.* Gen. Tech. Rep. RMRS-GTR-126. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 111 p. Available online: http://www.fs.fed.us/rm/pubs/rmrs_gtr126.html

Schumm, S.A. 1991. *To Interpret the Earth: Ten Ways to be Wrong.* Cambridge University Press. 133 p.

Schwing, F.B., T. Murphree, and P.M. Green. 2002. The Northern Oscillation Index (NOI): a new climate index for the northeast Pacific. *Progress in Oceanography* 53:115-139

Schwing, F.B., R. Mendelssohn, S.J. Bograd, J.E. Overland, M. Wang, and S. Ito. 2009. Climate change, teleconnection patterns, and regional processes forcing marine populations in the Pacific. *J. Mar. Syst.* doi:10.1016/j.jmarsys.2008.11.027

Scott, A. and C. Wild. 1991. Transformations and R^2. *The American Statistician* 45:127-129

Şen, Z., and Z. Habib. 2000. Spatial precipitation assessment with elevation by using point cumulative semivariogram technique. *Water Resources Management* 14:311-325. doi 10.1023/A:1008191012044

Sokal, R.R., and F.J. Rohlf. 1994. *Biometry*. 3rd ed. W.H. Freeman. 880 p.

Steinskog, D.J., D.B. Tjøstheim, and N.G. Kvamstø. 2007. A cautionary note on the use of the Kolmogorov–Smirnov Test for normality. *Monthly Weather Review* 135:1151–1157

Stenseth, N.C. and A. Mysterud. 2005. Weather packages: finding the right scale and composition of climate in ecology. *Journal of Animal Ecology* 74:1195–1198 doi: 10.1111/j.1365-2656.2005.01005.x

Stenseth, N. C., A. Mysterud, G. Ottersen, J.W. Hurrell, K.-S. Chan, and M. Lima. 2002. Ecological effects of climate fluctuations. *Science* 297:1292–1296

Stenseth, N.C., G. Ottersen, J.W. Hurrell, A. Mysterud, M. Lima, , K.-S. Chan, N.G. Yoccoz, and B. Ådlandsvik. 2003. Studying climate effects on ecology through the use of climate indices: the North Atlantic Oscillation, El Niño Southern Oscillation and beyond. *Proceedings of the Royal Society of London, Series B* 270:2087–2096

Stephenson, A., and E. Gilleland. 2005. Software for the analysis of extreme events: the current state and future directions. *Extremes* 8:87-109

Stephenson, D. 2002. Estimation and attribution of changes in extreme weather and climate events. In: *IPCC Workshop on Extreme Weather and Climate Events*, 11-13 June 2002, Beijing. Presentation. http://www.met.rdg.ac.uk/~han/Extremes/extreme1.pdf

Steward, R.R. 2005. The oceanic influence on North American drought. Chapter in: *Our Ocean Planet – Oceanography in the 21st Century.* Texas A&M University. On-line textbook: http://oceanworld.tamu.edu/resources/oceanography-book/oceananddrought.html

Stohlgren, T.J., T.N. Chase, R.A. Pielke, Sr., T.G.F. Kittel, and J. Baron. 1998. Evidence that local land use practices influence regional climate, vegetation, and stream flow patterns in adjacent natural areas. *Global Change Biology* 4:495-504

Thompson, D.W.J., and J.M. Wallace. 2000. Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate* 13:1000-1016

Thornton, P.E., H. Hasenauer, and M.A. White. 2000. Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application of complex terrain in Austria. *Agricultural and Forest Meteorology* 104:255-271

Toews, M.W., P.H. Whitfield, and D.M. Allen. 2007. Seasonal statistics: The 'seas' package for R. *Computers and Geosciences* 33:944-951

Tomé, A.R., and P.M.A. Miranda. 2004. Piecewise linear fitting and trend changing points of climate parameters. *Geophys. Res. Lett.* 31:L02207. doi:10.1029/2003GL019100

Torrence, C., and G.P. Compo. 1998. A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.* 79:61–78

Trenberth, K.E. 1984. Signal versus noise in the Southern Oscillation. *Monthly Weather Review* 112:326-332

Trenberth, K.E. 1997. The definition of El Niño. *Bull. Am. Meteorol. Soc.* 78:2771–2777

Trenberth, K.E. 1998. Atmospheric moisture residence times and cycling: Implications for rainfall rates with climate change. *Climatic Change* 39:667-694

Trenberth, K.E., and J.W. Hurrell. 1994. Decadal atmosphere-ocean variations in the Pacific. *Climate Dyn.* 9:303-319

Trenberth, K.E., A. Dai, R.M. Rasmussen, and D.B. Parsons. 2003. The changing character of precipitation. *Bull. Am. Meteorol. Soc.* 84:1205–1217

Trenberth, K.E., P.D. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. Klein Tank, D. Parker, F. Rahimzadeh, J.A. Renwick, M. Rusticucci, B. Soden and P. Zhai. 2007. Observations: surface and atmospheric climate change. In: Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.). *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, and New York, NY, USA.

Urban, D.L., R.V. O'Neill, and H.H. Shugart. 1987. Landscape ecology. *Bioscience* 37:119-127

Verzani, J. 2004. *Using R for Introductory Statistics.* Chapman & Hall/CRC. 432 p. In part on Google Books: http://books.google.com/books?id=jwolc192c5kC

Verosub, K.L., and J. Lippman. 2008. Global impacts of the 1600 eruption of Peru's Huaynaputina Volcano. *Eos* 89:141-142

von Storch, H., and F.W. Zwiers. 2001. *Statistical Analysis in Climate Research.* Cambridge University. 484 p. In part on Google Books: http://books.google.com.br/books?id=5QgAfL1N6koC

Vose, R.S., C.N. Williams Jr., T.C. Peterson, T.R. Karl, and D.R. Easterling. 2003. An evaluation of the time of observation bias adjustment in the US Historical Climatology Network. *Geophysical Research Letters* 30(20):2046, pp. CLM3-1 to 3-4. doi:10.1029/2003GL018111.

Wade, C.G. 1987. A quality control program for surface mesometeorological data. *J. Atmos. Oceanic Technol.* 4:435–453

Wallace, J.M., and D.S. Gutzler. 1981. Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.* 109:784–812

Wessa, P. 2009. *Free Statistics Software (v1.1.23-r1).* Office for Research Development and Education, Resa Corporation R&D: http://www.wessa.net/

Whitcher B., P. Guttorp, and D.B. Percival. 2000. Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research-Atmospheres* 105(D11):14941-14962

Wilks, D. 2006. *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press / Elsevier. Available on-line: http://www.scribd.com/doc/7128720/Statistical-Methods-in-the-Atmospheric-Sciences-Daniel-Wilks-

Willett, J. B., and J. D. Singer. 1988. Another cautionary note about R^2: its use in weighted least-squares regression analysis. *The American Statistician* 42:236-238

Willmott, C.J., C.M. Rowe, and W.D. Philpot. 1985. Small-scale climate maps: a sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. *The American Cartographer* 12:5-16

Wolter, K., and D. Allured. 2007. New climate divisions for monitoring and predicting climate in the U.S. *Intermountain West Climate Summary* 3(5):2-6 (June 2007). http://wwa.colorado.edu/IWCS/archive/IWCS_2007_Jun.pdf

Wolter, K., and M.S. Timlin. 1993. Monitoring ENSO in COADS with a seasonally adjusted principal component index. *Proc. of the 17th Climate Diagnostics Workshop*, Norman, OK, NOAA/N MC/CAC, NSSL, Oklahoma Clim. Survey, CIMMS and the School of Meteor., Univ. of Oklahoma, pp. 52-57

Wright, S.J., and O. Calderón. 2006. Seasonal, El Niño and longer term changes in flower and seed production in a moist tropical forest. *Ecology Letters* 9:35–44

Yandell, B.S. 1997. *Practical Data Analysis for Designed Experiments.* CRC Press. 437 p. In part on Google Books: http://books.google.com/books?id=K-e05Mrab0oC

Yang, D., D. Kane, Z. Zhang, D. Legates, and B. Goodison. 2005. Bias-corrections of long-term (1973-2004) daily precipitation data over the northern regions. *Geophysical Research Letters* 32:L19501, doi:10.1029/2005GL024057. see also: http://www.uaf.edu/water/faculty/yang/bcp/index.htm

Yiou, P., E. Baert, and M.F. Loutre. 1996. Spectral analysis of climate data. *Surveys in Geophysics* 17:619-663

You, J.S., K.G. Hubbard, S. Nadarajah, and K. Kunkel. 2007. Performance of quality assurance procedures on daily precipitation. *Journal of Atmospheric and Oceanic Technology* 24:821-834

Tables

Table 1. An example of a plausibility limits table that can be used in basic validity checks (after Burroughs 2008). Limits are set to catch values not physically plausible; these are tailored for a given site (§3.3.1). This is separate from screening for outliers which would employ tighter limits (§3.3.2).

Item	Valid Range
Year	1954 - present
Month	1 - 12
Day	1 - last day in corresponding month
Observation Hour	0 - 23
Temperature	-50 to +50 °C
Precipitation	0 to 100 mm/day
Dewpoint Depression	0 to 50 °C
Wind Speed	0 - 100 m/s
Wind Direction	0 - 360°
etc.	

Table 2. On-line resources for station metadata (indicated by an asterisk*; §3.4.1) and monitoring and outlook products providing near-real time, regional context to park climates (§4.7).

Source	Web Entry Point
*NOAA Regional Climate Centers	http://www.wrcc.dri.edu/rcc.html - links to all NOAA regional centers ▪ e.g., for Western U.S. – http://www.wrcc.dri.edu/CLIMATEDATA.html
*State Climatologists	http://www.stateclimate.org/ ▪ e.g., for Wyoming http://www.wrds.uwyo.edu/sco/climate_office.html
NOAA Climate Prediction Center	http://www.cpc.ncep.noaa.gov/ – Climate monitoring U.S., Pacific Islands, Global: ▪ http://www.cpc.ncep.noaa.gov/products/monitoring_and_data/ ▪ http://www.cpc.ncep.noaa.gov/products/precip/CWlink/
NOAA National Climate Data Center	http://www.ncdc.noaa.gov/ ▪ http://www.ncdc.noaa.gov/oa/climate/research/monitoring.html ▪ http://lwf.ncdc.noaa.gov/oa/climate/research/cag3/cag3.html
National Integrated Drought Information System	http://www.drought.gov/ ▪ National Drought Mitigation Center – http://drought.unl.edu/
Natural Resource Conservation Service	Snow course maps – http://www.wcc.nrcs.usda.gov/snowcourse/
Western Water Assessment	http://wwa.colorado.edu/IWCS/index.html - Intermountain West climate summary ▪ http://wwa.colorado.edu/forecasts_and_outlooks/forecasts.html - Links to other US climate-related websites

Table 3. Web entry-point resources for major teleconnection patterns for United States and territories. Sites give descriptions and maps for teleconnection patterns for annual and monthly surface climate variables; some sites also include teleconnections for extremes. Oscillation abbreviations are given in the text §4.8.1. Some patterns are illustrated in Figure 24 (for NPI, ENSO, PDO, and AMO). Data sources for oscillation indices are given in Table 4. A broader summary of Northern Hemisphere teleconnections is provided at: http://www.cpc.noaa.gov/data/teledoc/telecontents.shtml

Teleconnection Pattern*†	Web Resources
ENSO teleconnections	▪ http://www.cru.uea.ac.uk/cru/info/enso/ ▪ http://www.esrl.noaa.gov/psd/enso// – o http://www.esrl.noaa.gov/psd/enso//enso.climate.html o http://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/composites/ U.S. by climate region and state – o http://www.cpc.ncep.noaa.gov/products/monitoring_and_data/ENSO_connections.shtml U.S.: for climate extremes – o http://www.esrl.noaa.gov/psd/enso/climaterisks/ http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/lanina/us_impacts/ustp_impacts.shtml ▪ http://www.cses.washington.edu/cig/pnwc/aboutenso.shtml o See *PDO teleconnections* for 'combined PDO & ENSO effects' ▪ http://jisao.washington.edu/analyses0500/#enso
PDO teleconnections	▪ http://www.atmos.washington.edu/~mantua/REPORTS/PDO/PDO_cs.htm; also Mantua et al. (1997) ▪ http://www.cses.washington.edu/cig/pnwc/aboutpdo.shtml o Combined PDO & ENSO effects – ▪ http://www.cses.washington.edu/cig/pnwc/clvariability.shtml ▪ http://www.cses.washington.edu/cig/pnwc/compensopdo.shtml ▪ http://jisao.washington.edu/analyses0500/#pdo ▪ http://www.beringclimate.noaa.gov/data/BCinclude.php?filename=in_PDO ▪ See *AMO teleconnections* for 'combined AMO & PDO effects'
& the related North Pacific Index (NPI)–Pacific/North American (PNA) teleconnections	▪ NPI: see Trenberth and Hurrell (1994), Hurrell (1996) ▪ http://jisao.washington.edu/data/pna/ ▪ http://jisao.washington.edu/analyses0500/#pna
NAM/AO & NAO teleconnections	▪ http://jisao.washington.edu/analyses0500/#ao ▪ http://www.cpc.ncep.noaa.gov/data/teledoc/nao.shtml ▪ http://nsidc.org/arcticmet/patterns/arctic_oscillation.html ▪ http://www.ldeo.columbia.edu/res/pi/NAO/ ▪ http://www.cru.uea.ac.uk/cru/info/nao/
AMO teleconnections	▪ http://oceanworld.tamu.edu/resources/oceanography-book/oceananddrought.html o Includes combined AMO & PDO effects; also McCabe et al. (2004) ▪ http://www.aoml.noaa.gov/phod/amo_faq.php
Other Major Teleconnection Patterns – East Atlantic (EA) West Pacific (WP) East Pacific-North Pacific (EP-NP) Tropical/Northern Hemisphere (TNH) Pacific Transition (PT)	▪ http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml

* For creating custom teleconnection correlation maps, see: http://www.esrl.noaa.gov/psd/data/correlation/.[122]
† For a glossary of terms related to circulation oscillations and teleconnections, see http://www.ucar.edu/news/backgrounders/patterns.shtml

Table 4. Hemispheric circulation oscillations indices and their online data sources for oscillations with significant teleconnections affecting the United States and territories. Related oscillations are listed together. The analysis variable that underlies each index is given in [brackets]. Several of the indices are plotted in Figure 23 (PDO, SOI, and MEI). Links describing their US teleconnection patterns are listed in Table 3. A complementary compilation of index sources is at: http://www.esrl.noaa.gov/psd/data/climateindices/list/. (Link to text §4.8.1 and 4.8.2)

Oscillation	Circulation Indices [Analysis Variable]*	Sources for Index Timeseries**
El Niño / Southern Oscillation (ENSO)	Niño Region 3.4 SST Index / Oceanic Niño Index (ONI)[152] [SST]	http://www.cgd.ucar.edu/cas/catalog/climind/Nino_3_3.4_indices.html http://www.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/enso.shtml ▪ http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml
	Southern Oscillation Index (SOI) [SLP]	http://www.cgd.ucar.edu/cas/catalog/climind/soi.html http://www.cru.uea.ac.uk/cru/data/soi.htm
	Multivariate ENSO Index (MEI) [multivariate]*	http://www.esrl.noaa.gov/psd/people/klaus.wolter/MEI/mei.html ▪ http://www.esrl.noaa.gov/psd/people/klaus.wolter/MEI/table.html
Pacific Decadal Oscillation (PDO)[130]	PDO Index [SST]	http://jisao.washington.edu/pdo/ ▪ http://jisao.washington.edu/pdo/PDO.latest [154]
	North Pacific Index (NPI)[153] [SLP]	http://www.cgd.ucar.edu/cas/jhurrell/npindex.html
North Atlantic Oscillation (NAO)[131]	NAO Index [SLP]	http://www.cgd.ucar.edu/cas/jhurrell/indices.html [156] http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml http://www.cru.uea.ac.uk/~timo/projpages/nao_update.htm
Northern Annular Mode (NAM) / Arctic Oscillation (AO)[155]	NAM Index [SLP]	http://www.cgd.ucar.edu/cas/jhurrell/indices.info.html#nam
	AO Index [1000mb ht]	http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao.shtml
Atlantic Multidecadal Oscillation (AMO)	AMO Index [SST]	http://www.esrl.noaa.gov/psd/data/timeseries/AMO/

* SST = sea surface temperatures, SLP = sea level pressure, ht = pressure surface height. MEI is based on SLP, SST, surface wind, surface air temperature, and cloudiness anomalies.
** Where more than one source is listed, it is generally because they use slightly different means to determine the index. Note that some sites present both monthly and seasonal series for some indices. I've attempted to list sites that keep index series up-to-date, but no guarantee.

[152] Trenberth (1997) explains the difference between these similar indices based on Niño Region 3.4 SST's: http://www.cgd.ucar.edu/cas/catalog/climind/. SST's in this region of the equatorial Pacific is a strong indicator of ENSO events, especially with respect to global teleconnections: http://www.ucar.edu/news/backgrounders/patterns.shtml#mno

[153] The North Pacific Index (NPI) reflects the PDO – not the second NPO pattern discussed in footnote 129 (Minobe and Mantua 1999). Note that the NPI is negatively correlated with the PDO index.

[154] SST source data for this PDO Index changed in 1982 and 2002 – see data source notes (and graphic link at end of file).

[155] See discussion of Arctic/Antarctic Annular Modes at http://ao.atmos.colostate.edu/introduction.html and their indices at http://ao.atmos.colostate.edu/Data/index.html

[156] Hurrell lists an array of NAO indices, for a overview see: http://www.cgd.ucar.edu/cas/jhurrell/naointro.html

Figures[157]

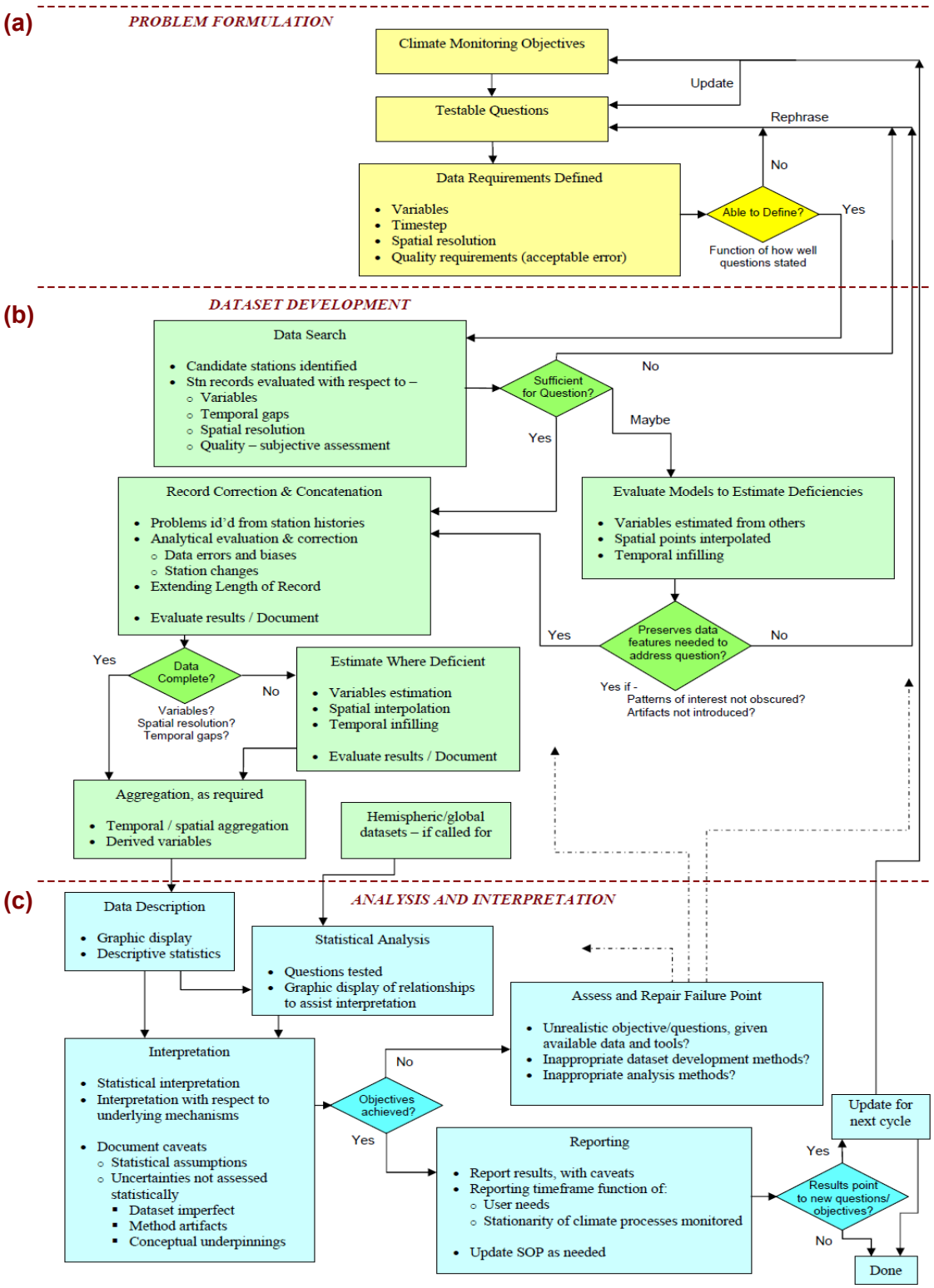

Figure 1. Workflow diagram laying out the overall process for developing and analyzing climate datasets. Three key stages are: (a) problem formulation (covered in section §2.0), (b) dataset development (§3.0), and (c) analysis and results interpretation (§4.0). Specific objectives and questions will dictate specific procedures for each stage. Not all boxes are covered in this report, such as variable estimation ('Evaluate Models' box), reporting, and program updating processes. (SOP = standard operating procedure document.) (Link to text §1.2)

[157] *N.B.* Many figures are from copyrighted publications. Permissions pending.

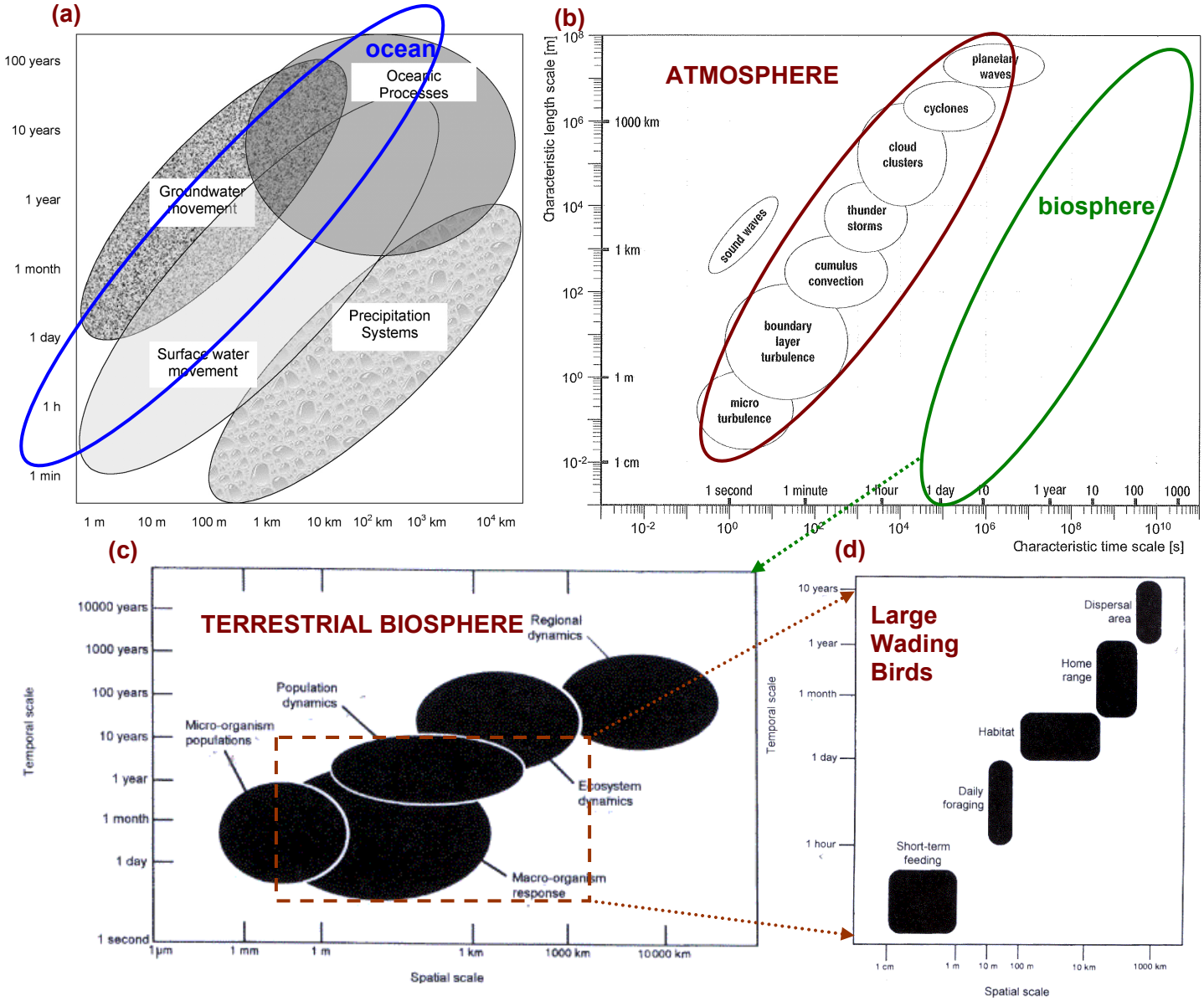

Figure 2. The relationship between characteristic temporal and spatial scales for (a) various geophysical processes in general,158 (b) atmospheric processes in particular (from von Storch and Zwiers 2001), and (c,d) terrestrial ecological processes.159 The blue oval in (a) roughly gives an alternate scheme for oceans from von Storch and Zwiers's (2001); the green oval in (b) generally represents that for the terrestrial biosphere in (c). Panel (d) gives a specific example for the behavior of large wading birds [the domain of this figure is indicated by the brown box in (c)]. Note that temporal and spatial axes are log10, and those in (b) are switched from those in (a,c,d). (in a-d, color annotations added)

Discussion. Regardless of the geophysical or ecological system, these figures show a positive log-log relationship between characteristic temporal and spatial scales for biogeophysical processes: larger processes take longer to operate. How this relationship lays out is, however, system dependent, with little or no overlap in the time-space relationship for climate and ecological processes (b). This means there is not a simple translation of variation in one system to the other at a given scale, but rather that there must be scale interactions. Differences among time-space relationships are illustrated in detail for ecological external forcings vs. internal processes by Urban et al. (1987; see also Delcourt et al. 1983). (Link to text §2.2)

---

158 Image from Water Cycle Study Group, 2001, Predictability of Variations in Global and Regional Water Cycles, Ch. 3, in: *A Plan for a New Science Initiative on the Global Water Cycle*. USGCRP. (http://www.usgcrp.gov/usgcrp/Library/watercycle/wcsgreport2001/wcsg2001chapter3.htm)

159 Image from Sheehan, P., 1995, Assessments of Ecological Impacts on a Regional Scale, Ch. 14 in: *SCOPE 53 - Methods to Assess the Effects of Chemicals On Ecosystems* (http://www.icsu-scope.org/downloadpubs/scope53/chapter14.html)

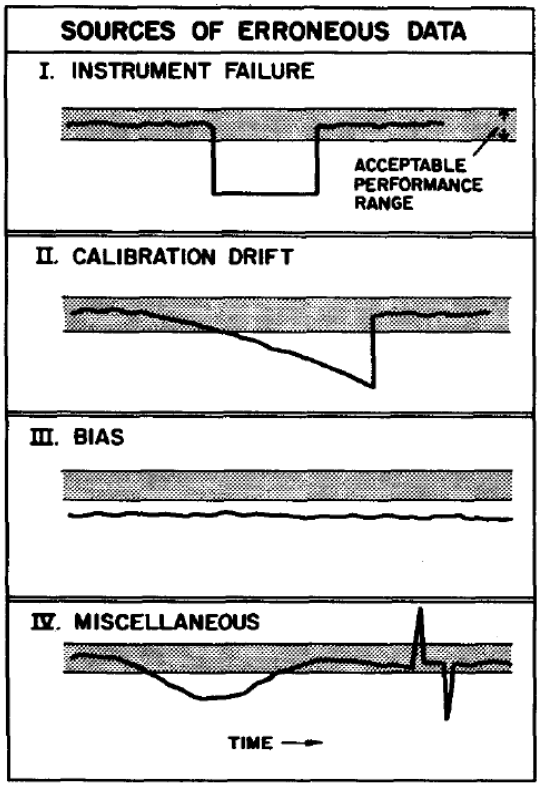

Figure 3. Types of errors often present in data from electronic sensors and other automated instruments (from Wade 1987).

Discussion: When errors carry the data outside of an expected dynamic range, they can be detected by straightforward screening techniques (e.g., plausibility tests, §3.3.1). Detection and correction of more subtle errors (with physically reasonable values), require more involved techniques (§3.3.2–3.3.3).

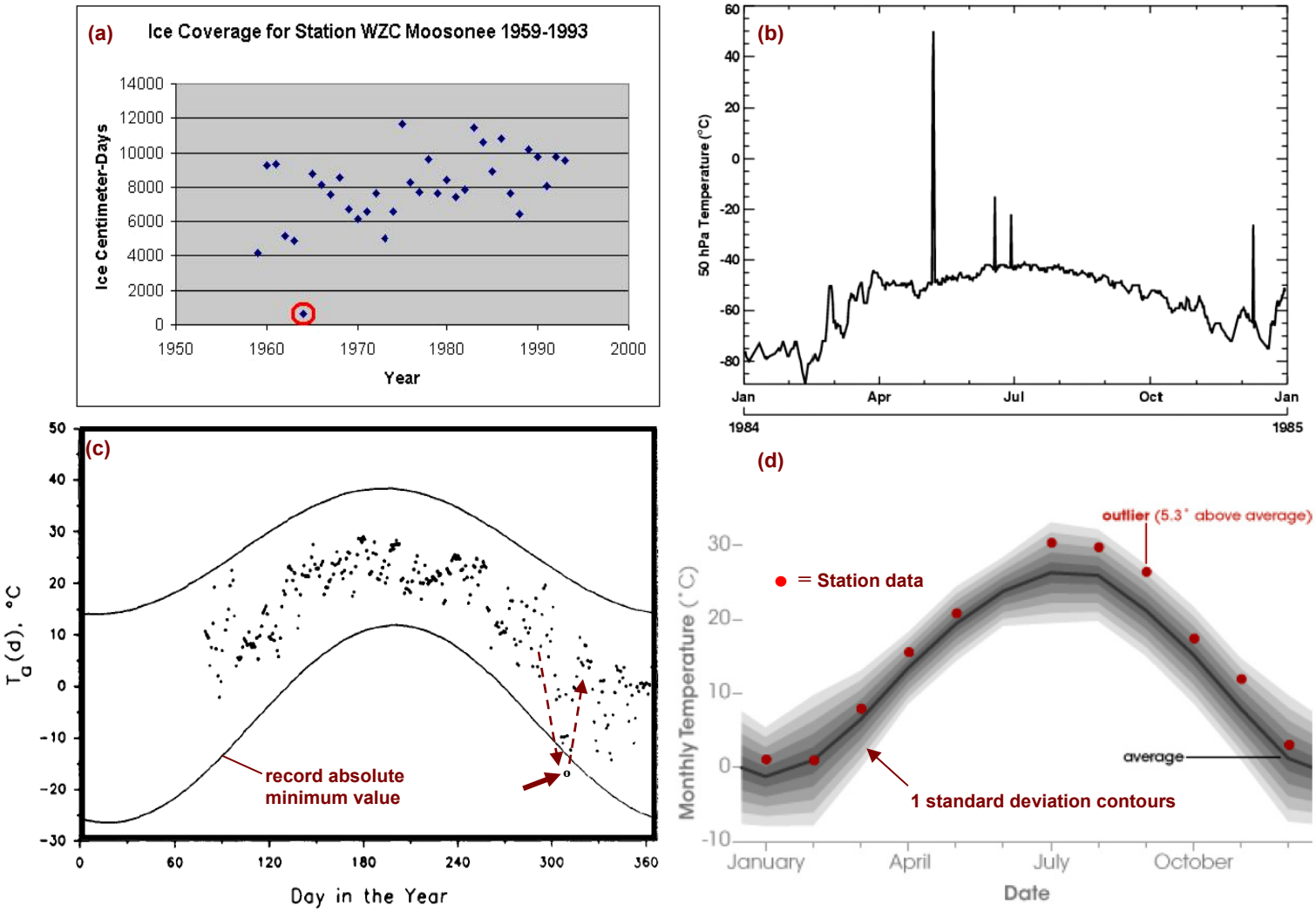

Figure 4. Illustrations of record outliers and different detection methods. (a) Ice coverage for a lake in Ontario (from Drews 2003). The circled outlier was manually detected. While physically plausible, the outlier was rejected when a manual review of observations found that the record for the suspect year was incomplete. (b) Tropospheric temperature record over Norway in 1984. The outliers (spikes) passed physical plausibility tests but were detected and rejected using a temporal consistency test (from Burroughs 2008). (c) Daily mean surface air temperature for Walnut Creek, CA. Automated screening used day-dependent outlier detection limits (curves) based on record absolute daily maxima and minima – this identified a late fall outlier (indicated by the solid arrow). However, the method can be overly sensitive as it does not allow any leeway for new valid extremes. Follow-on visual inspection showed that the outlier was temporally consistent with a week-long cold snap (dashed arrows show the trace of dailies leading down to and then up from the outlier) and so was not rejected (from Meek and Hatfield 1994; brown annotations added). (d) The 1932 monthly temperature record for Linyi, China (dots) was automatically screened first with climatological limits based on 2.5 and 5 standard deviations (SD) from the longterm mean seasonal cycle (black curve; gray-shaded contours are 1 SD increments from the average). Depending on which SD threshold was exceeded, outliers were further evaluated for spatial consistency against observations from nearby stations and for temporal consistency. Finally, flagged outliers were plotted for visual checks – if the datapoint was not clearly faulty, it was retained. In this case, the September 1932 mean was identified as outlier (> 2.5 SD), but was confirmed as reasonable value by other stations' records and was retained. (Hansen et al. 1999; figure from Herring 2007; brown annotations added) (Link to text §3.3.2)

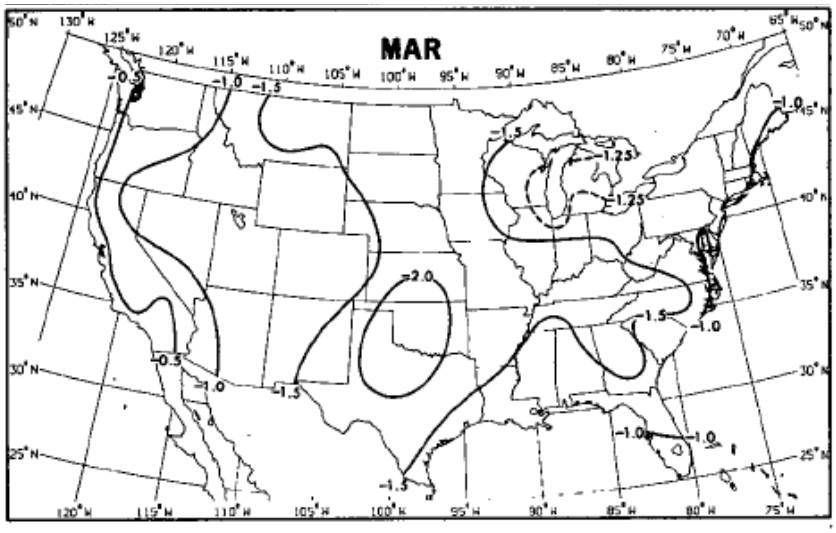

Figure 5. Time-of-observation bias. Difference in March mean monthly temperature between daily minimum/maximum thermometer observations taken in the early morning vs. end of the afternoon (0700h vs. 1700h local time). (from Karl et al. 1986)

Discussion: Temperature bias due to time of observation is strongest for continental regions and in spring (as shown here) and late fall. This geographic and seasonal pattern of high biases is related to the prevalence in the interior and in spring and fall of strong warm and cold fronts, which are linked to this bias (see text §3.3.3). Morning-based records give monthly means consistently colder than afternoon ones by as much as 2ºC or greater. Over the last 50 years, U.S. reporting times have steadily shifted from late afternoon to early morning, contributing an artificial cooling tendency to uncorrected regional trends (Vose et al. 2003).

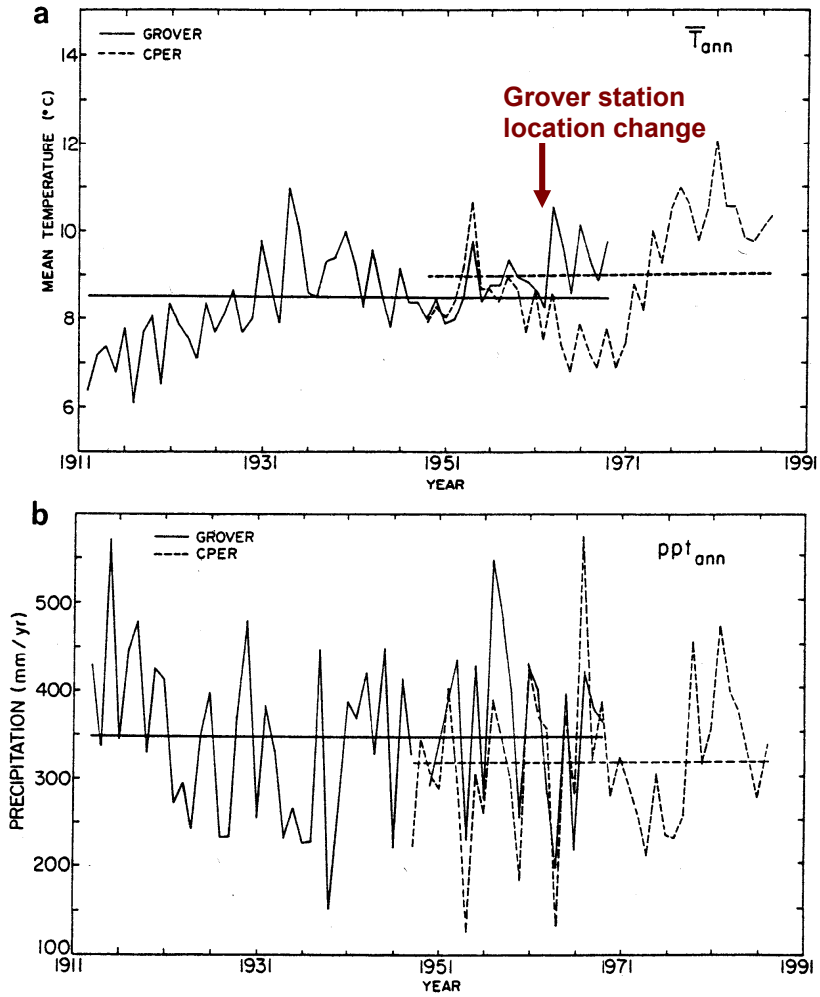

Figure 6. Comparing records of nearby stations, illustrating issues with combining records. Water-year (Oct-Sept) (a) mean temperature and (b) precipitation for two stations in the Colorado shortgrass steppe: Grover, CO and the Central Plains Experimental Range (CPER) LTER (Long-Term Ecological Research Program) site. (from Kittel 1990; in a, brown annotation added)

Discussion: The objective was to create a combined record for exploring longterm climate dynamics of the region. Three issues must be addressed: (1) station differences in means, (2) station differences in variance structure, and (3) degree of correlation in interannual variability. For the period of overlap, the two sites track each other well until 1962 (correlation $r = +0.71$ for temperature, +0.77 for precipitation, $p<0.01$). This correlation breaks down after 1962 (arrow) especially for temperature (entire overlap period 1949-69: $r = +0.14$, n.s.). A check of the metadata for Grover revealed that the station was moved a significant distance in 1962. A truncated overlap period (prior to the Grover station change, 1949-1962) was then used to evaluate station differences in annual means and standard deviations, which were not grossly different. The long overlap permitted spectral analysis, which also revealed that variance structure was similar at sub- and multidecadal scales. These results confirmed that the annual records could be combined (concatenated) by adjusting records based on differences in means and ratios of standard deviations. (Link to text §3.5).

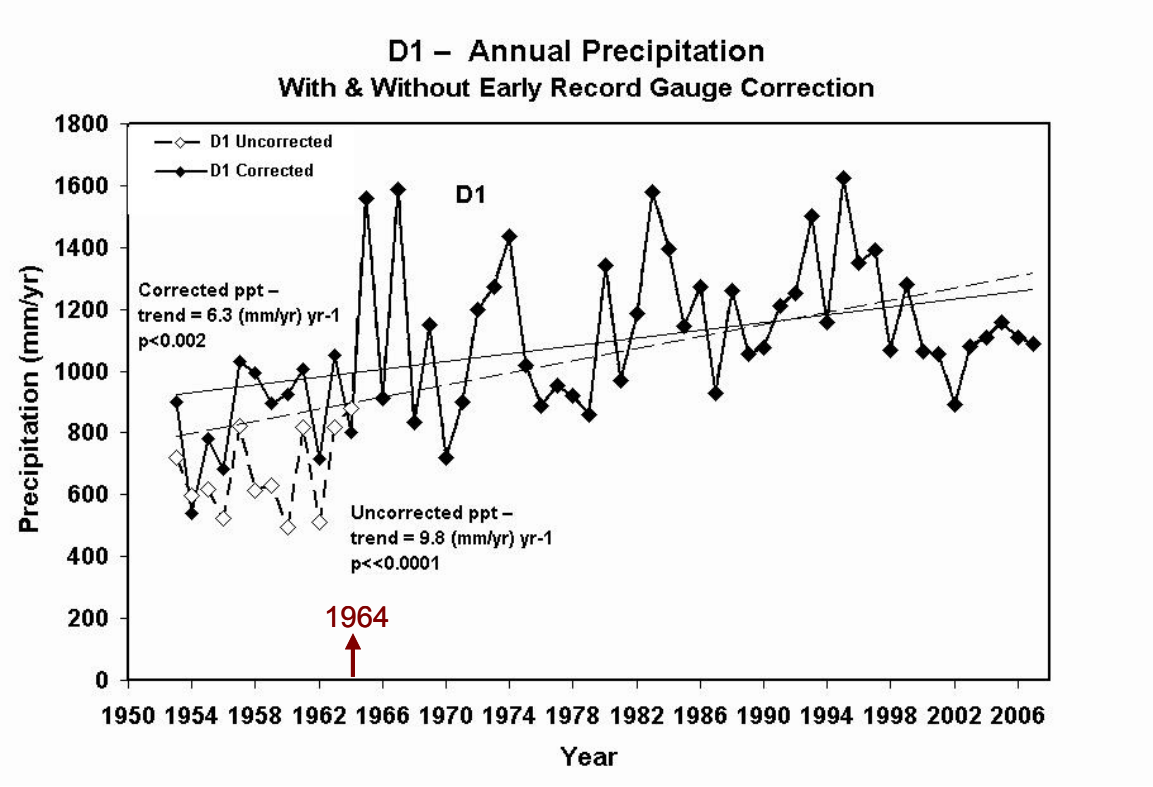

Figure 7. Accounting for a precipitation gauge change at a high elevation station in the Colorado Front Range. A unshielded Belford gauge was replaced at the end of 1964 with a recording bucket gauge along with an Alter-type shield and a Wyoming fence to prevent wind-caused undercatch. A 2-year overlap in instruments permitted the development of a correction factor, applied to the original data (*dashed line*) and reflected in the solid line (through 1964). (Kittel et al., in preparation)

Discussion: The correction had a marked effect on the longterm trend analysis: with correction, the trend decreased from +9.8 mm yr^{-1} yr^{-1} ($p<10^{-5}$) to +6.3 mm yr^{-1} yr^{-1} ($p<0.002$). (Link to text §3.4)

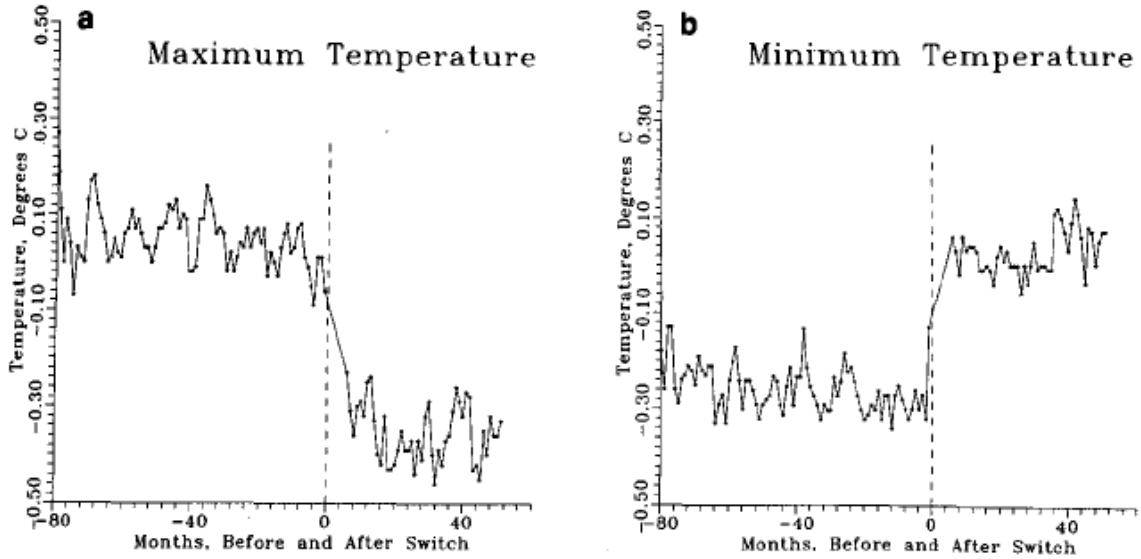

Figure 8. Step changes in (a) maximum and (b) minimum temperature records in the mid-1980's due to a change from liquid-in-glass thermometers to the electronic sensor Maximum/Minimum Temperature System (MMTS) (from Quayle et al. 1991). Plotted values are the average timeseries across stations in the conterminous U.S., co-registered to the time of the change over.

Discussion: Opposing step changes in T_{min} and T_{max} resulted in a severely reduced diurnal temperature range after the switch. Such instrument artifacts severely interfere with climate trend assessments. Quayle et al. (1991) presents bias correction factors to adjust records for this change. (Link to text §3.4.5)

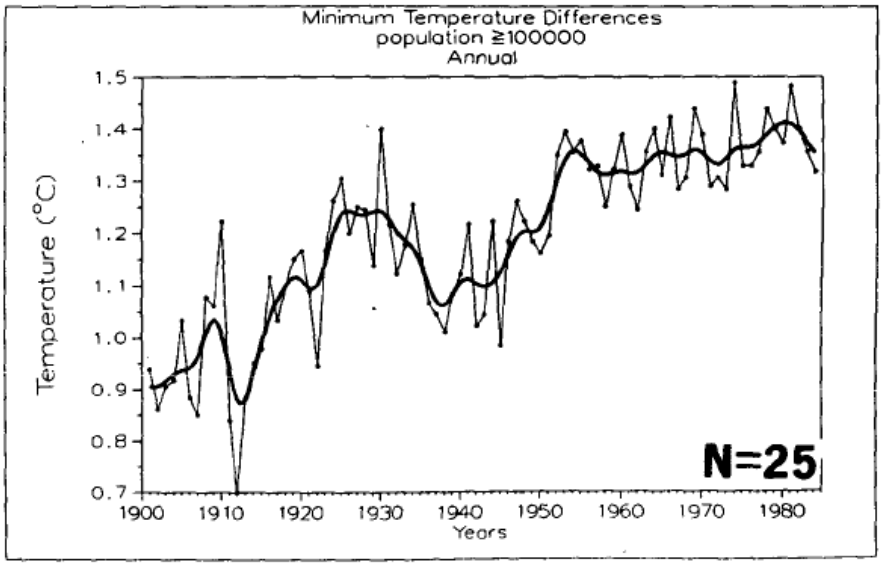

Figure 9. Urban heat island effect: Evolution of differences in annual minimum temperature between stations in cities (1980 populations ≥100,000) relative to paired rural stations (populations <2,000) (N= number of pairs). (from Karl et al. 1988)

Discussion: The magnitude of the urban heating trend is substantial (~0.5ºC/80y), especially when compared to non-urban influenced global trends. The urban effect is real, but is often removed from regional datasets designed to evaluate trends arising from other sources (e.g., the USHCN dataset[17,59]). (Link to text §3.4.7)

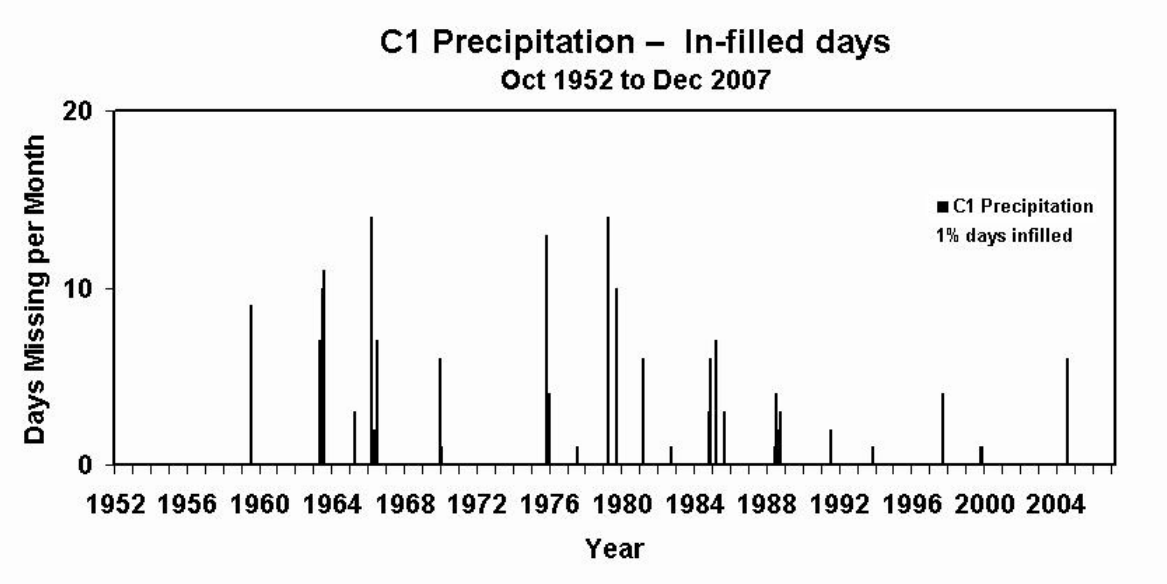

Figure 10. Number of days missing per month in the precipitation record of a high elevation station in the Colorado Front Range (Niwot Ridge). (Kittel et al., in preparation)

Discussion: The plot shows that the record is nearly complete with missing values well distributed – in particular, not concentrated at either end of the record which otherwise would raise the question of temporal biasing (§3.3.1). Missing values were infilled using adjacent instrument and nearby station records (§3.6).

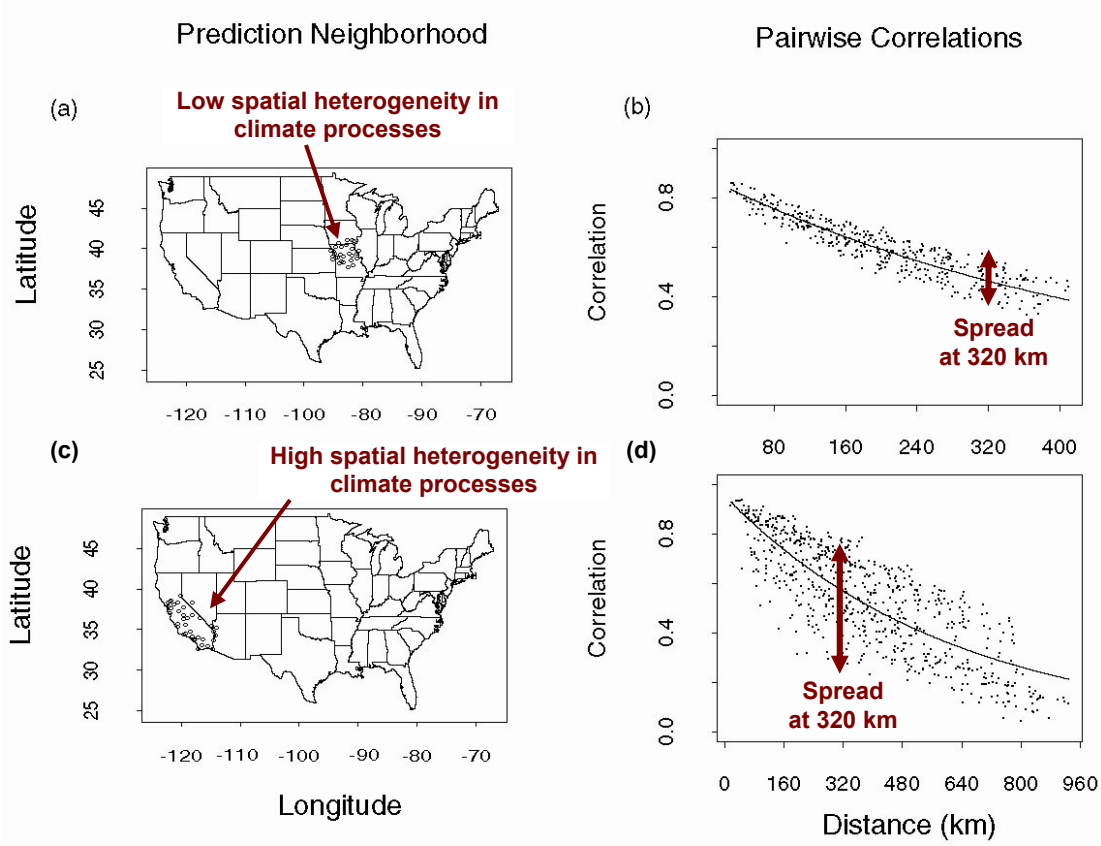

Figure 11. Spatial autocorrelation structure in monthly precipitation anomalies is dependent on heterogeneity in factors controlling climate, demonstrated here for two climate regions: (a, b) in the Midwest where autocorrelation structure was strong and (c, d) in southern California where it was weak. (a, c) Maps show a neighborhood ('local window') of 20 stations used to develop site correlograms shown in (b, d). In (b, d), note that *x*-axis distance scales differ between these two figures (vertical double-headed arrows are placed at *x*=320 km to facilitate comparison). Site correlograms show correlations among all station pairs as a function of distance (dots), along with the best fit line from an exponential model. (from Kittel et al. 2004; brown annotations added). (Link to text: §3.6.4)

Discussion: The double arrows show that the spread of correlations at a distance of 320 km was greater for a more climatically heterogeneous California than for the more uniform Midwest.

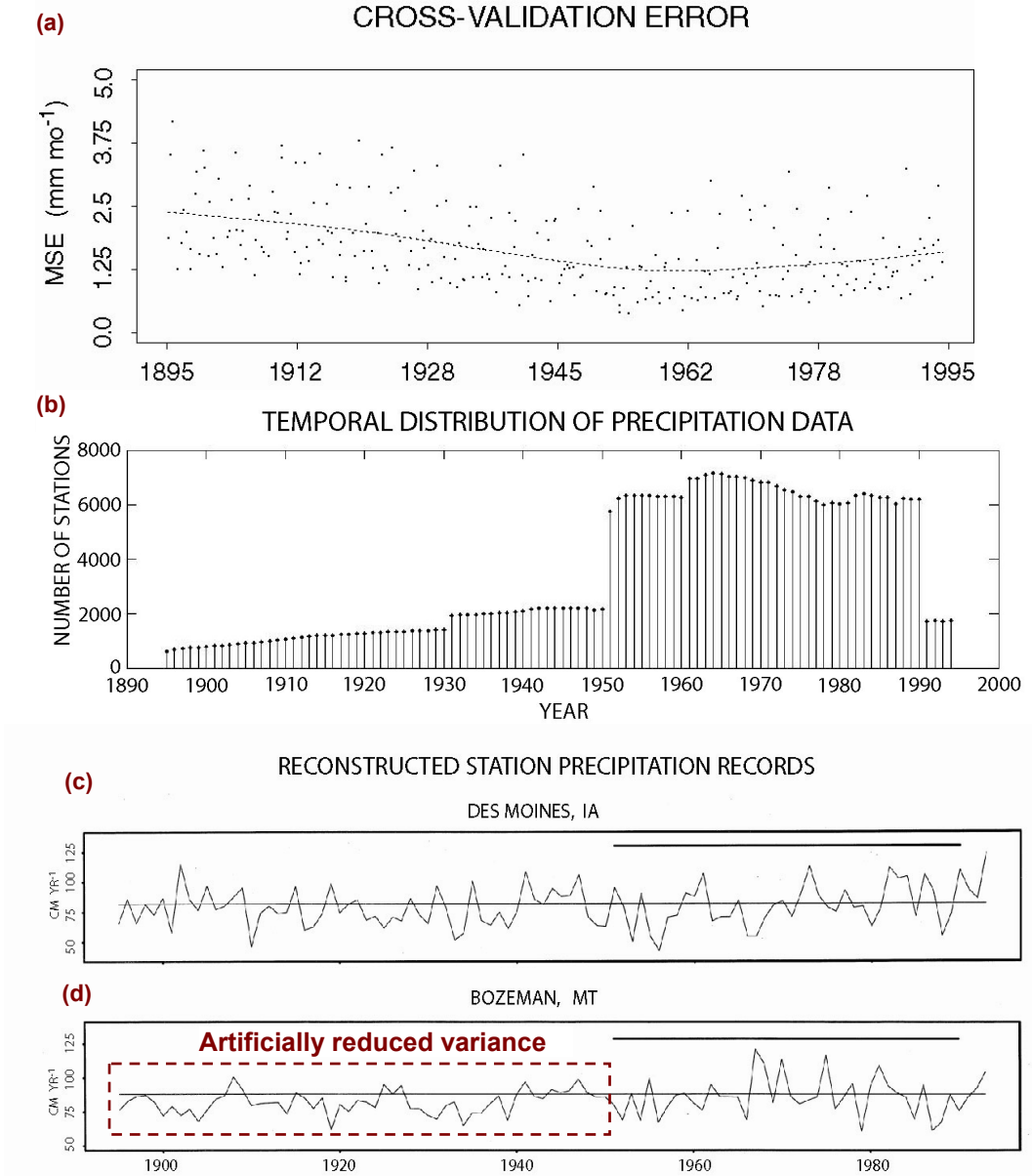

Figure 12. Time and station-density dependence in errors for the spatial (kriging) model in Figure 11. (a) Time dependence in cross-validation errors for precipitation, expressed as mean squared error (MSE) for square-root transformed precipitation, is higher during periods of reduced station density in (b). Cross-validation is where the model predicts a subset of observations withheld from the original analysis. The decrease in station number after 1990 was due to delayed updating of some national datasets. (c,d) Visualization of the effects of station density on interannual variability of infilled station precipitation records created using the kriging model, for different climatic regions: (c) Midwest (Des Moines) and (d) the Mountain West (Bozeman). The heavy horizontal bar in the upper right spans the period when observed data were available for the station. (from Kittel et al. 2004; in d, brown annotation added)

Discussion: Note in (a) that the model is relatively robust to station density change: average precipitation errors roughly doubled from recent decades back to the early part of the record (MSE = 1.3 to 2.5 mm/mo), corresponding to a nearly 10-fold decrease in station numbers across the domain (in b). In (d), interannual variability was artificially reduced during the early part of reconstructed records for areas where station densities were depleted, as around Bozeman. This is a consequence of the model generating overly smoothed spatial fields when there were too few station observations to adequately capture regional variability. As station density drops, the process reached farther away from a site to find predictor stations and blended unrelated anomaly patterns from adjacent regions to make a point's prediction. Poorly-related anomalies tended to counter each other, diminishing overall variance in reconstructed timeseries. This is not an issue with infilled series for regions where stations densities were sufficiently high early in the record, for example, in the Midwest (in c). (Link to text §3.6.5)

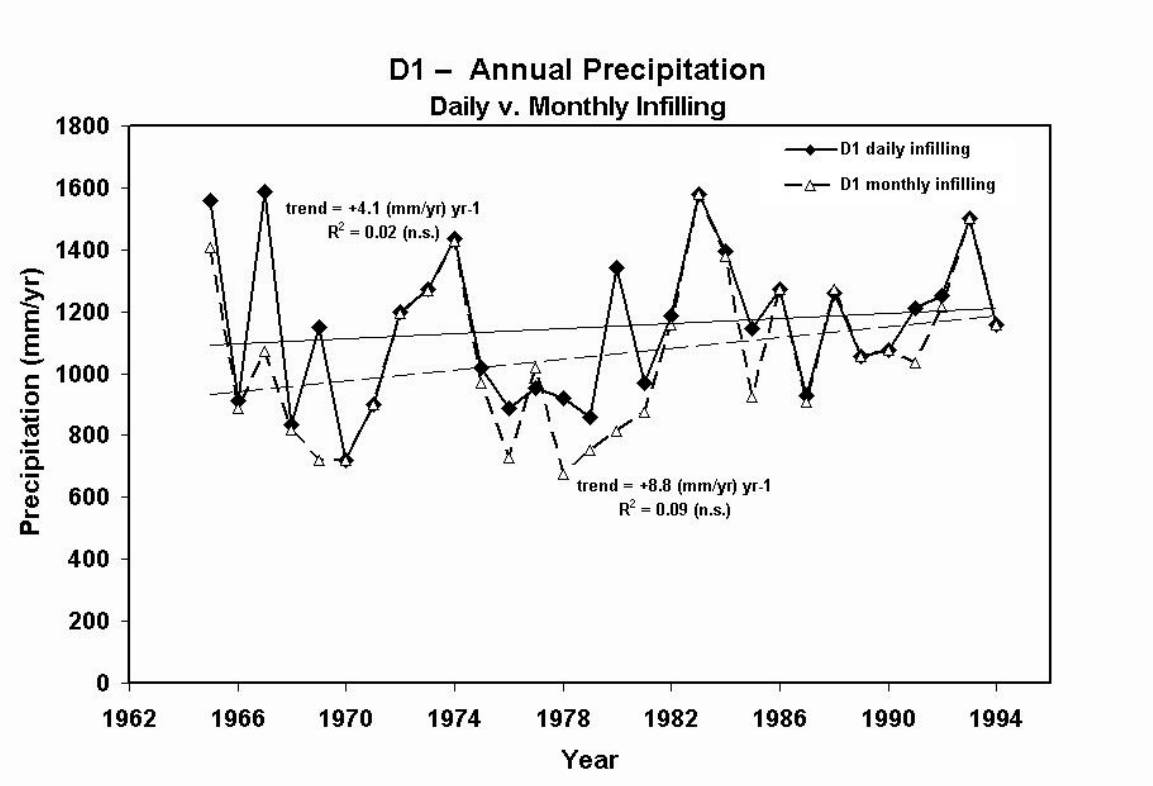

Figure 13. The effects of applying a missing-value method at different temporal resolutions on the annual precipitation record at a high elevation station in the Colorado Front Range (Figure 7, Figure 10) for 1965-1999. The long-standing technique was to infill monthly precipitation for months with missing days using monthly correlations with nearby stations (*dashed curve*). Later, the same spatial method was repeated but applied at a daily timestep based on daily correlations (*solid curve*). (Kittel et al., in preparation)

Discussion: While infilling method did not change the significance of longterm trends (both nonsignificant, n.s.), the daily method produced a markedly different series. Differences in infilled data were accumulated over the year and were due to the daily model being able to select different stations for the regression on a daily basis, rather than just one to represent a month. The daily-method timeseries was taken to be more true to the station record as it preserved more of the observed data. (Link to text §3.6.5)

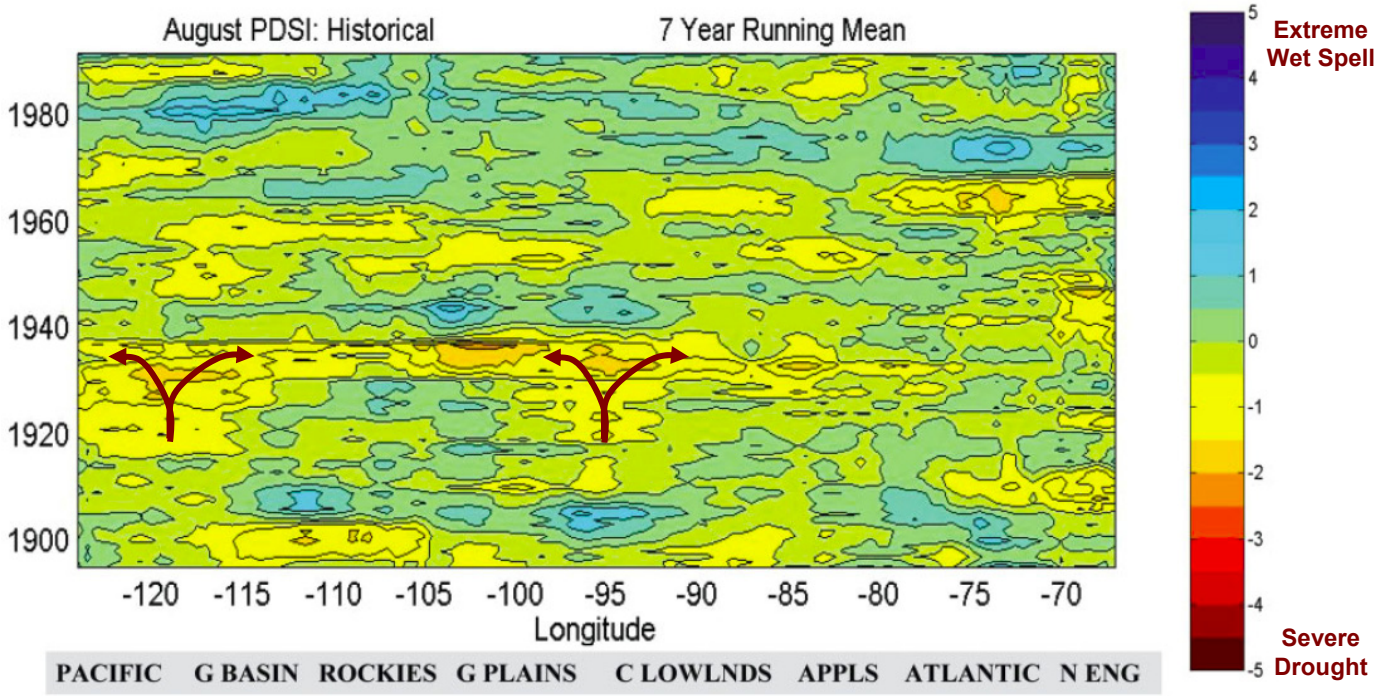

Figure 14. A longitude-time section plot[72] of the Palmer Drought Severity Index (PDSI) for August for the conterminous United States (from Kittel et al. 2004; brown annotations added). Vertical axis is time (year), horizontal axis is longitude (°W). Mapped values are latitudinal (north-south) averages for a given year and longitude, and smoothed using a 7-year running average. Longitude labels: Pacific = Pacific states, G Basin = Great Basin, Rockies = Rocky Mountains, G Plains = Great Plains, C Lowlnds = Central Lowlnds, Appls = Appalachians, Atlantic = Atlantic Coastal Plain, N Eng = New England.

Discussion: The longitude-time section plot shows how droughts (yellow-deep orange) in the far western and central US developed in the 1920's, intensified and merged in the 1930's, and then rapidly dissipated. (Links to text: section plots §4.1.2, PDSI §4.2.2)

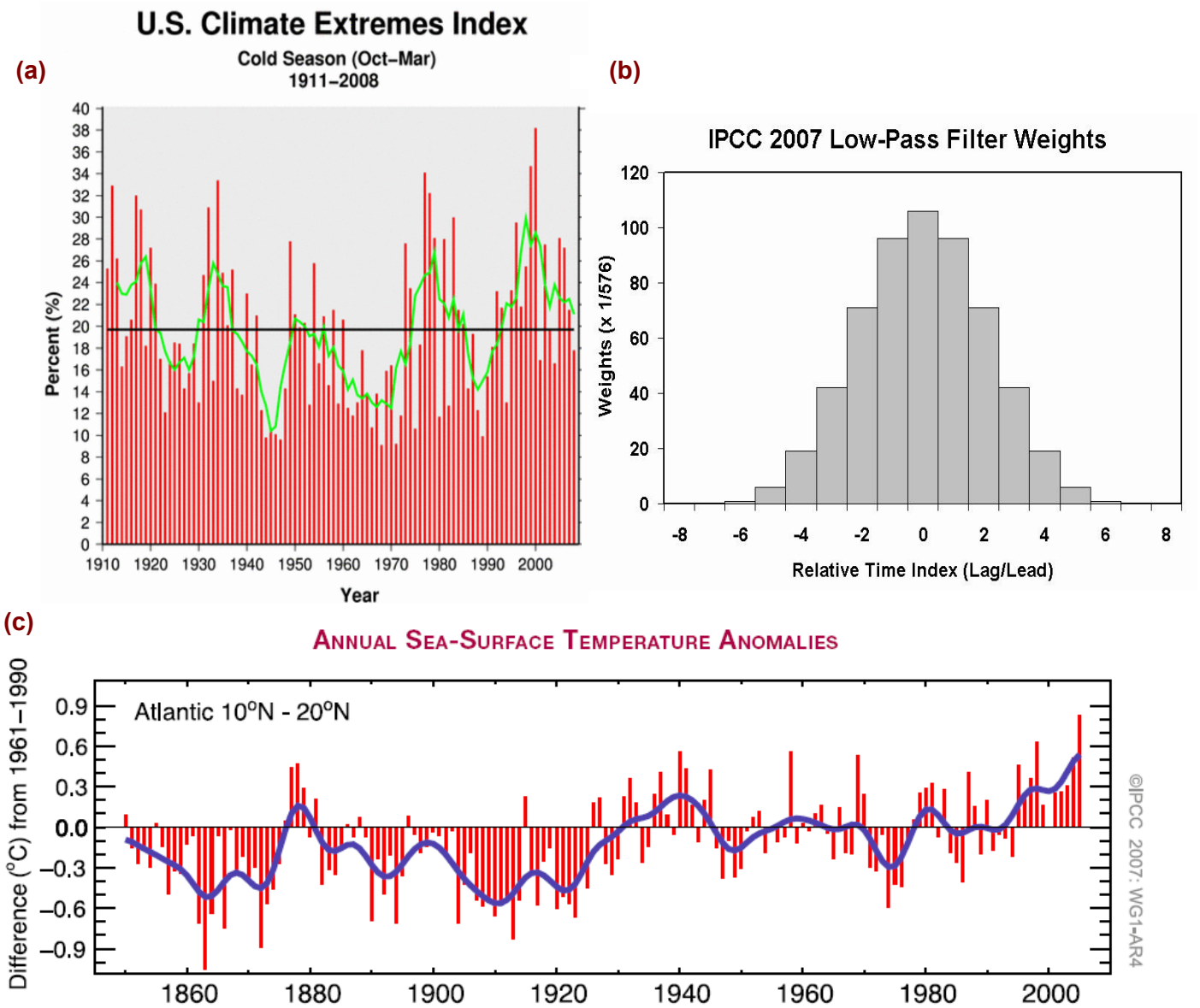

Figure 15. Temporal smoothing revealing decadal and longer-period oscillatory behavior. (a) Timeseries of the pervasiveness of cold-season climate extremes across the conterminous United States, as represented by the aggregate Climate Extremes Index (CEI, from Gleason et al. 2008; §4.6.2).[160] Bars are yearly values, smoothed curve is a 5-year centered moving average, and black horizontal line is the period average. The value of CEI is the percent of the three dimensional time-space domain (season × region) experiencing extreme climate. The occurrence of extreme climate is based on any of five thermal and moisture measures exceeding their upper or lower 10%-tiles. (b) Weights used in IPCC WG1's centered low-pass 13-year moving filter (figure created from data in Trenberth et al. 2007). The 13 weights have numerators {1-6-19-42-71-96-106-96-71-42-19-6-1} and a denominator of 576. (c) 1850-2005 tropical Atlantic sea surface temperature annual anomalies (bars) and smoothed curve (thick line) using weights in (b) designed to remove less than decadal scale fluctuations (from Trenberth et al. 2007). At the end points, where 13 points are not longer available to the filter, the smoothed line is constrained by minimizing its slope (Mann 2004; see text §4.5.1).

Discussion: In (a), the smoothed timeline suggests that the cold-season extremes index varies with a period of roughly 20 years. In (c), the low-pass filter curve suggests several scales of quasi-periodic behavior with decadal and longer periods. In both cases (a, c), the next step would be to evaluate the original series using spectral techniques (§4.5.2). (Link to text §4.5.1)

[160] http://www.ncdc.noaa.gov/oa/climate/research/cei/cei.html

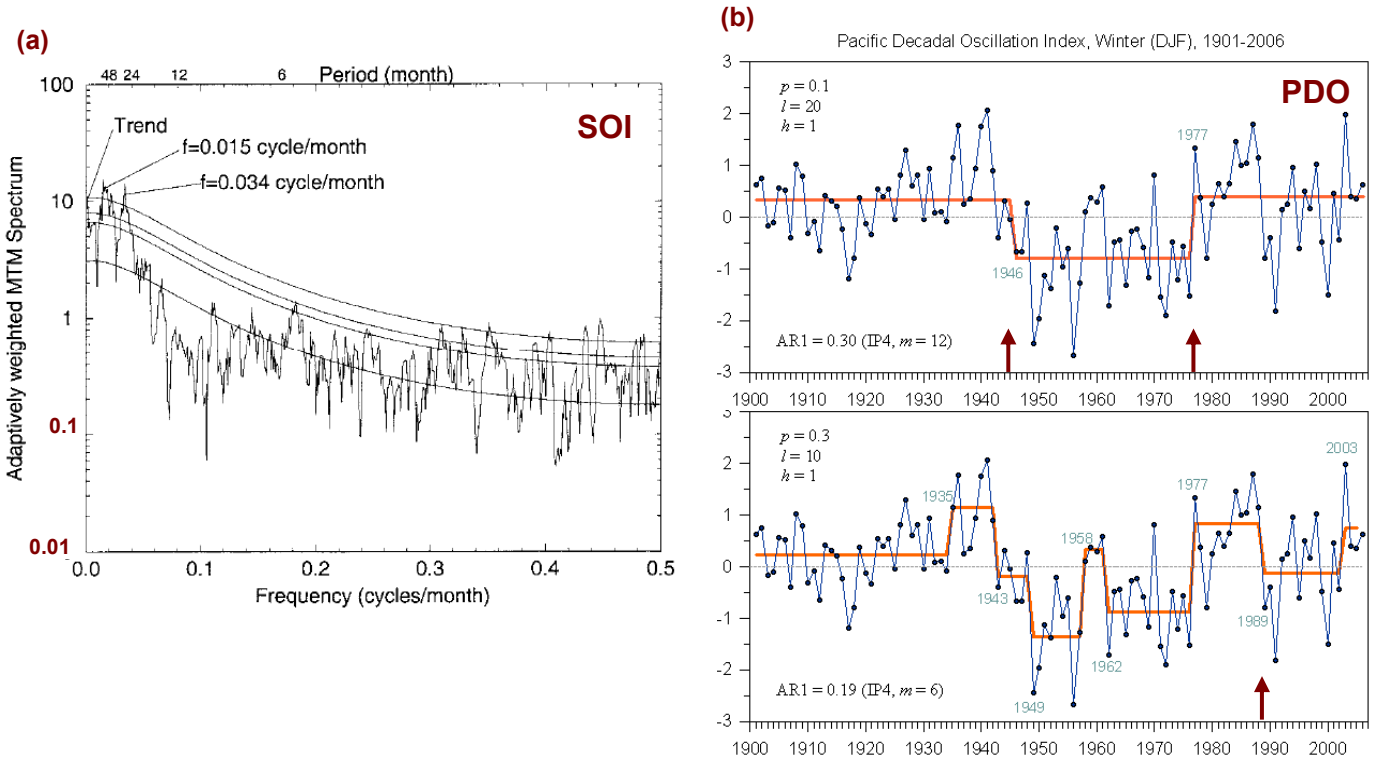

Figure 16. Multiyear temporal pattern analysis. (a) Spectral analysis of the Southern Oscillation Index (SOI; §4.8, Figure 23b) (from Ghil et al. 2002; brown annotations added). Spectral power is plotted on the *y*-axis. The bottom *x*-axis is frequency (1/month), converted to period (months) along the top *x*-axis. The uppermost smoothed line is the 99% confidence level, and below that, the 95% confidence level. (b) Regime shift detection in the Pacific Decadal Oscillation (PDO; §4.8, Figure 23a) for winter (from Rodionov and Overland 2005; brown annotations added).[161] *Top* vs. *bottom panels* show a sensitivity to detection model parameters

Discussion: At interannual or lower frequencies [left end of the spectrum in (a)], three peaks exceed the 99% level. Those marked 0.015 and 0.034 cycle/month correspond to dominant modes of the SOI, with periods spanning 2-6 years. The third, left-most peak is the longterm trend. (Link to text §4.5.2) In (b: *top*), two major regime shifts are detected in North Pacific climate dynamics over the last century: in 1946 and more recently in 1977 (marked by vertical arrows). Smaller, less persistent shifts were detected with relaxed cutoff-length thresholds in the regime shift analysis, including a warm-to-cold shift in 1989 (b: *bottom*). These two recent shifts (1977, 1989) are consistent with shifts in an array of abiotic and biotic measures across the northeastern Pacific (Hare and Mantua 2000; see also §4.8.1, §4.8.2). (Link to text §4.5.3)

[161] Image from: http://www.beringclimate.noaa.gov/data/BCinclude.php?filename=in_PDO. Model parameters: *p* target significance level, *l* cutoff length (years), *h* Huber weight parameter, and AR1 autoregressive parameter are noted in links to specific images and described in Rodionov and Overland (2005).

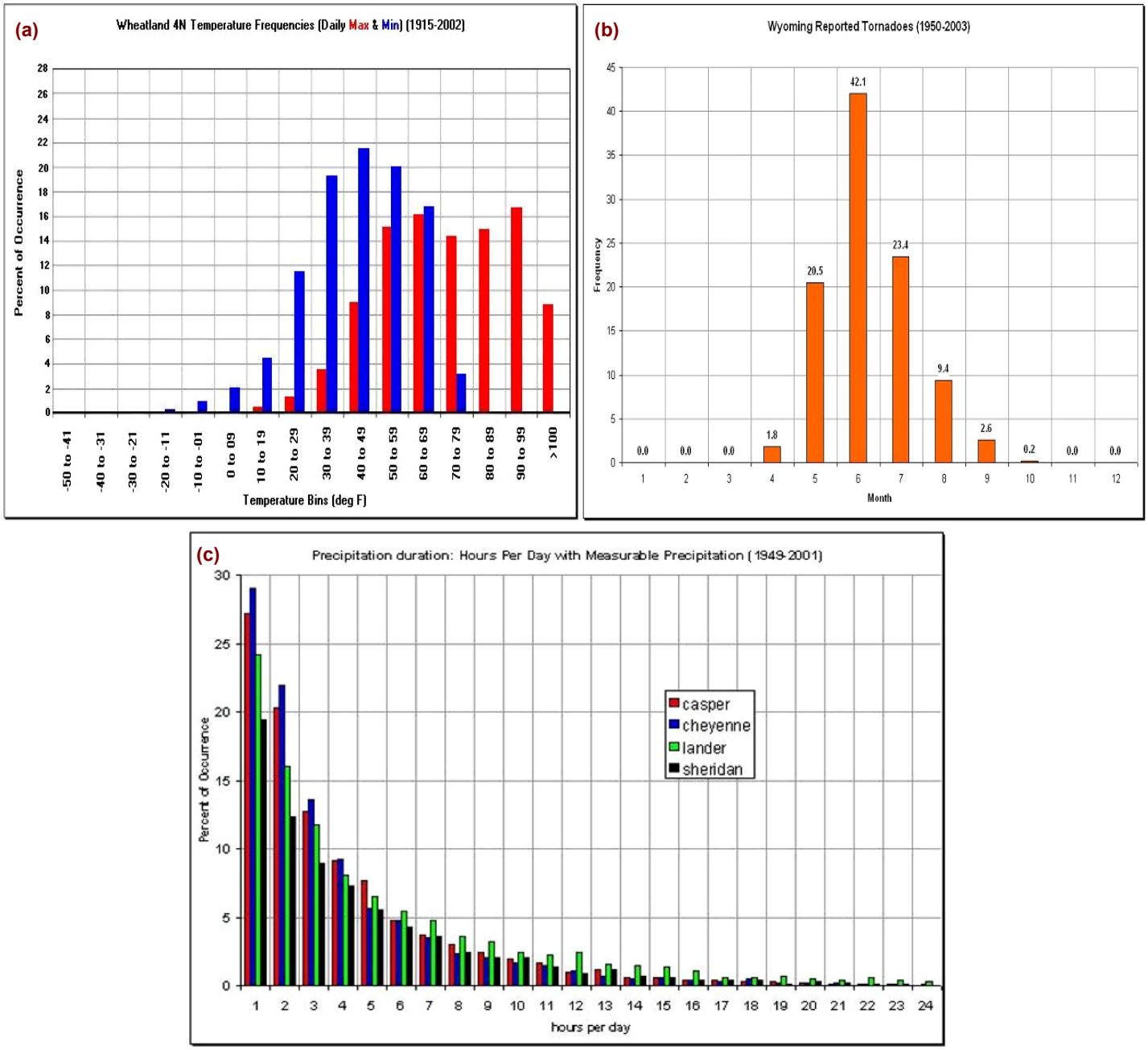

Figure 17. Evaluating the frequency of events. Frequency distribution of (a) daily T_{min} and T_{max} for Wheatland, WY by 10ºF (~6ºC) bins, (b) tornados in Wyoming by month of the year, and (c) precipitation duration by hours per day bins for four stations in Wyoming.[162]

Discussion: The occurrence of these three kinds of weather events is evaluated by binning by event magnitude (a, c) or timing (b). In (a) minimum temperatures (blue bars) at Wheatland are slightly negatively skewed, with a stretched out distribution of the most extreme cold events. Maximum temperatures (red bars) are more broadly distributed, but note that the bins are truncated at the high end (>100ºF), so the distribution of extreme maxima cannot be evaluated. Here, there is a suggestion of bimodal structure, which might reflect two separate (perhaps seasonal) processes, warranting further analysis. In (b), the seasonal distribution of tornados is slightly positively skewed, with more occurrences loaded up at the beginning of the warm season. This distribution is consistent with our understanding that strong gradients in temperature and/or moisture – most often occurring in late spring through early summer – are precursors for tornado development. For (c), as is characteristic for dry climates, precipitation events are highly concentrated on the short-duration end and are primarily less than one hour, yet with some full-day (or perhaps longer) events. Cheyenne appears to have more very short events, Lander more longer events – this difficult could be tested with a Kolmogorov-Smirnov identical distribution test (§4.6.3). (Link to text §4.6.1)

[162] Images from Wyoming Climate Atlas, http://www.wrds.uwyo.edu/sco/climateatlas/title_page.html, 17 Oct 2008 update.

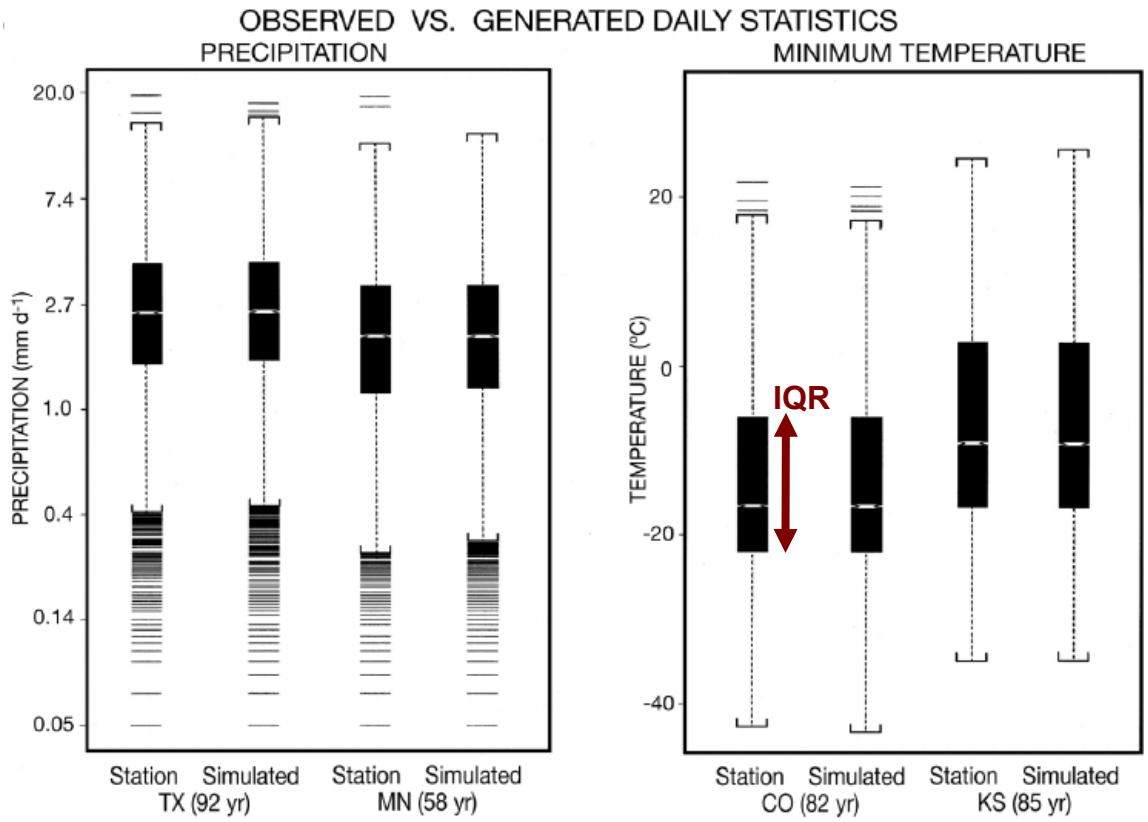

Figure 18. Structure of daily weather events as illustrated by notch plots. Plots are for January (*left panel*) precipitation and (*right*) minimum temperature from daily station records and daily weather model simulations for selected points (from Kittel et al. 2004; brown annotation added). The daily weather model was WGEN (§3.6.7).[69] Points are in central Texas (TX), southern Minnesota (MN), western Colorado (CO), and northwestern Kansas (KS). Record length (years) is indicated in parentheses after each location. Notch plots show median (center of notch), standard deviation (white area = 2 SD), interquartile range (IQR; dark boxes around median, spanning 25 to 75% of values), data range (brackets), and outliers (bars beyond the brackets; defined as values > +1.5 IQR or < – 1.5 IQR). Precipitation is for wet days (only non-zero values are included) and is plotted in natural-log space (though note, *y*-axis tick labels are in original precipitation units, mm/day). Box plots present the same distribution information but show the median without the standard deviation 'notch.'

Discussion: This graphic analysis shows that daily precipitation event size distribution is similar for both Texas and Minnesota stations with the median size and distribution set slightly higher at the Texas site. The distributions are symmetrical in natural-log space, meaning in linear space that the distributions have long tails on the high end. Both sites are characterized by many small (≪0.4 mm/day) outlier events. The daily minimum temperatures distribution for the Colorado and Kansas stations are similar, both asymmetrical with a broader distribution above the median than below. Relative to the Kansas site, the Colorado station has a distribution overall shifted towards lower minimum temperatures, a less broad IQR, and a few high outliers over the 82-year record. (Link to text §4.6) The frequency distribution of WGEN-simulated values strongly follows observed (§3.6.7).

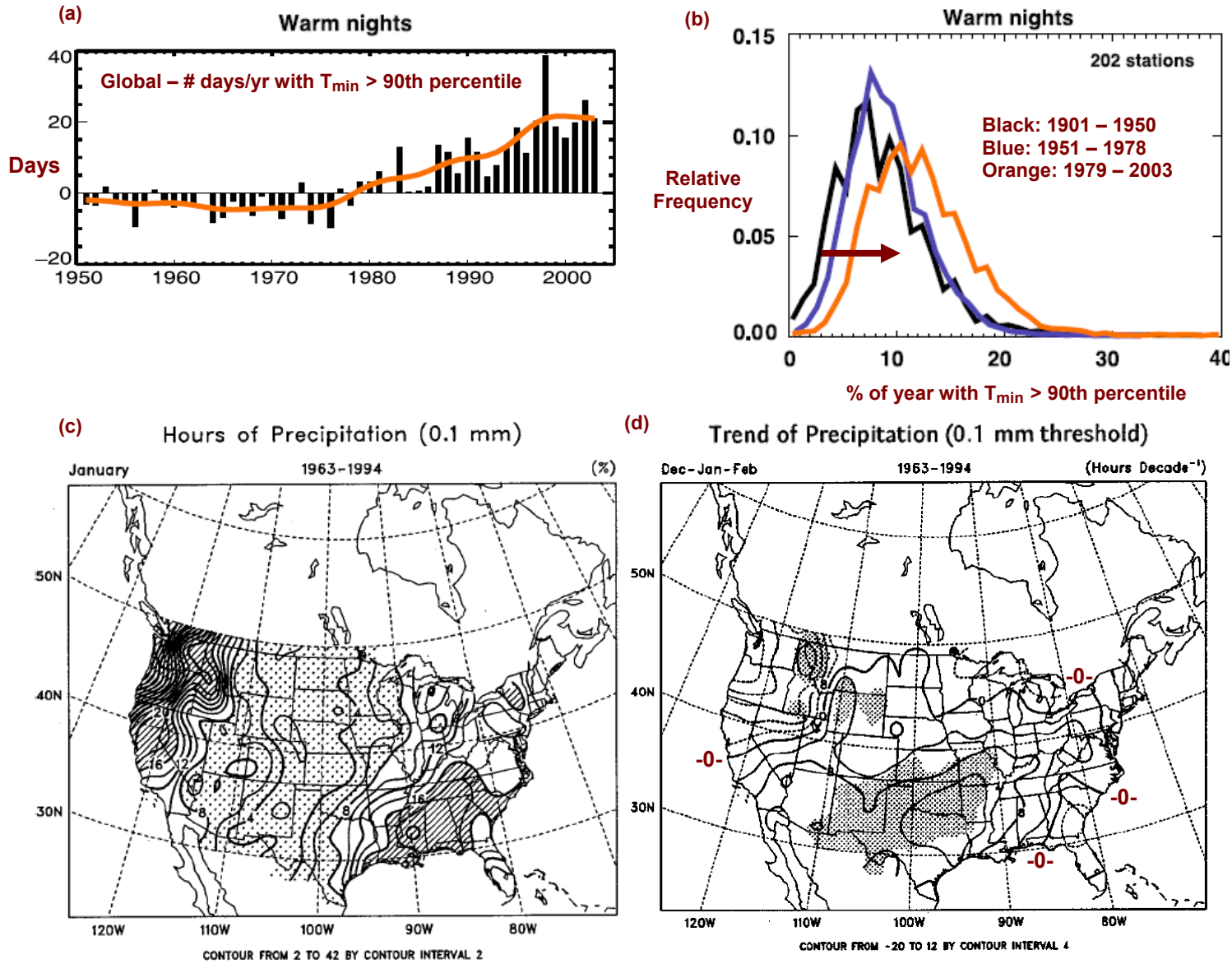

Figure 19. Historical changes in temperature and precipitation event frequency. (a, b) Global change in the occurrence of high extreme daily T_{min} ('warm nights'), defined as T_{min}'s in the top 10% of observations: (a) days with warm nights by year and (b) the frequency distribution of the percent-of-year with warm nights based on 202 stations, evaluated for 3 successive multidecadal periods (from Trenberth et al. 2007, based on Alexander et al. 2006; brown annotations added). Smoothed line in (a) is a 21-term filter.[163] (c, d) Frequency of precipitation over the conterminous United States (for all but the smallest events, ≥ 0.1 mm/hour) (from Trenberth 1998): (c) January precipitation frequency (percentage of hours with precipitation) and (d) historical trend in winter precipitation frequency (change in precipitation hours per decade) over the period 1963–1994. In (c), areas with less than 8% of hours are stippled and greater than 16% are hatched (contour interval=2%). In (d), negative trends are shown with dashed contour lines (in the Southeast and Pacific Northwest) and zero or positive change with solid contours (contour interval=4 hours/decade; brown -0- contour labels added). Areas with trends that are statistically significant (at 5% level) are stippled.

Discussion: In (a) – on a global basis, the number of days with warm nighttime temperature extremes increased from the early part of the 20th century to the beginning of the present one. The resulting greater portion of the year experiencing warm nights is expressed in (b) in terms of both a *positive shift* (horizontal arrow) in and a *broadening* of the frequency distribution of the % of year with warm nights. The next step would be to evaluate changes in the frequency timeseries (a) using trend analysis (§4.3), while changes in the frequency distribution (b) could be analyzed using *identical-distribution tests* (§4.6.3). Winter U.S. precipitation frequency [shown for January in (c)] is greatest in the Southeast and Pacific Northwest and low throughout the continental interior. (d) shows statistically significant late 20th century changes in this pattern, with precipitation frequency reduced most strongly in the northern U.S. Rockies and enhanced broadly across the Southern Great Plains and into the Southwest. (Link to text §4.6)

[163] http://hadobs.metoffice.com/hadcrut3/smoothing.html

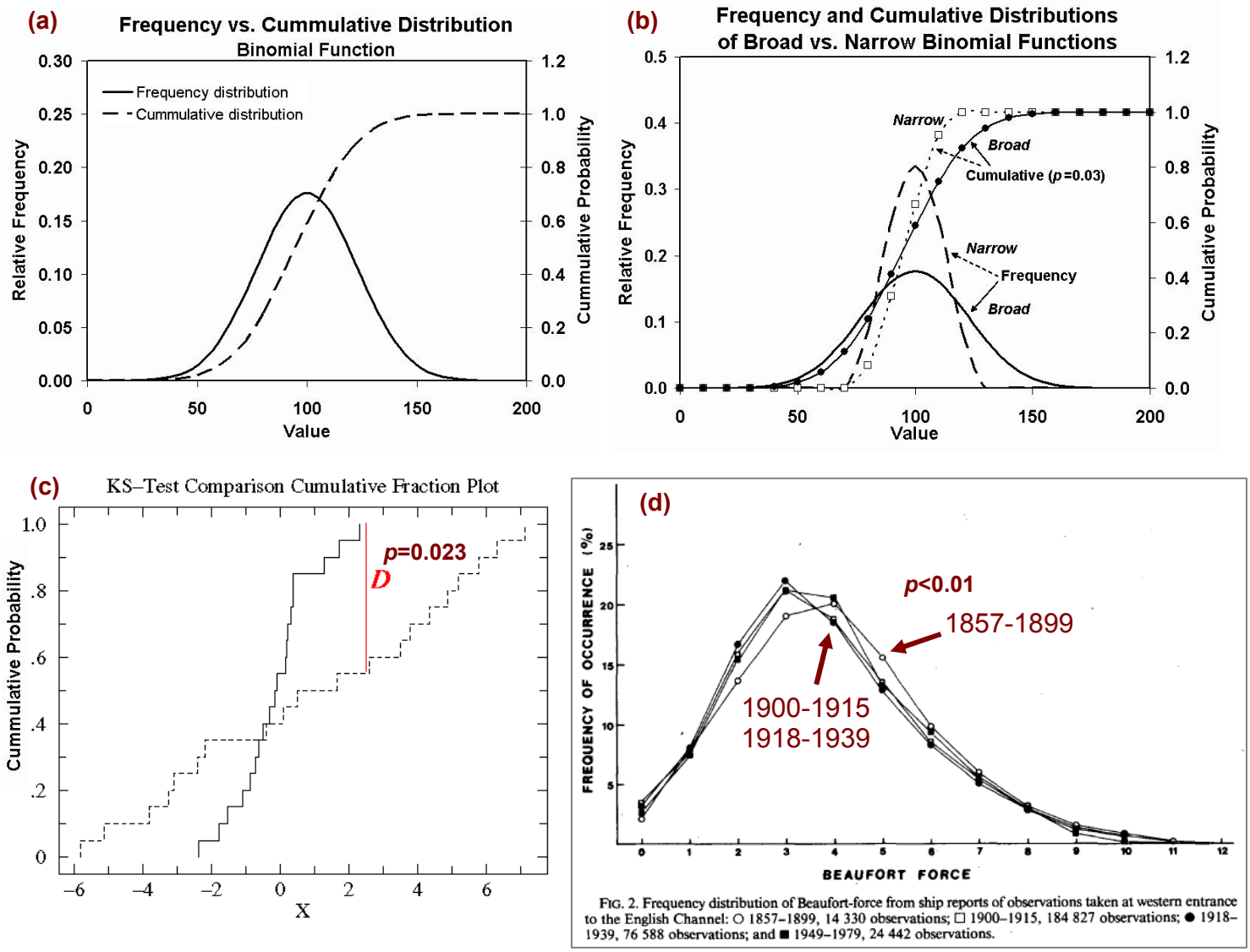

Figure 20. (a) Cumulative probability function (dashed line; right *x*-axis scale) for a binomial relative frequency distribution function (solid line; left axis). (b) Comparison of two theoretical binomial curves with narrow (dashed lines) vs. broad (solid lines) dispersion, plotted as both relative frequency (simple lines) and cumulative (lines with symbols) functions. The cumulative probability curves are significantly different (p=0.03, Kolmogorov-Smirnov identical distribution test, 'KS-test' – evaluated with an online calculator[115]). (c) Comparison of empirical cumulative functions for two sets of observations with similar means (t-test, nonsignificant at p>>0.05) but with strongly differing variances, resulting in statistically different distributions (KS-test, p=0.023). D is the maximum vertical deviation between the two curves and is the Kolmogorov-Smirnov test statistic (from Kirkman 1996; brown annotation added).[164] (d) Comparison of Beaufort force shipboard wind records in the English Channel for successive multidecadal periods. The frequency distribution of observations pre-1900 (open circles) was significantly different from subsequent records through 1939 (open squares and closed circles) (KS-test, p<0.01) (from Peterson and Hasse 1987; brown annotation added).

Discussion: The two cumulative distributions in (c) have similar means (at x=0). However, just as in the theoretical example (b), the dispersion of observations in (c) about the mean is much stronger for one set of observations (dashed curve, which starts earlier and climbs more slowly) than the other (solid curve). The significance of this difference is confirmed with the KS-test. In (d), the KS-test allowed Peterson and Hasse (1987) to identify a significant distribution shift in the Beaufort force record after 1899 – they suggest this maybe an artifact of a shift from sailing to steam ships at the end of the 19th century, perhaps related to concomitant changes in navigation and observer habits. (Link to text §4.6.3)

[164] Image from: http://www.physics.csbsju.edu/stats/KS-test.html

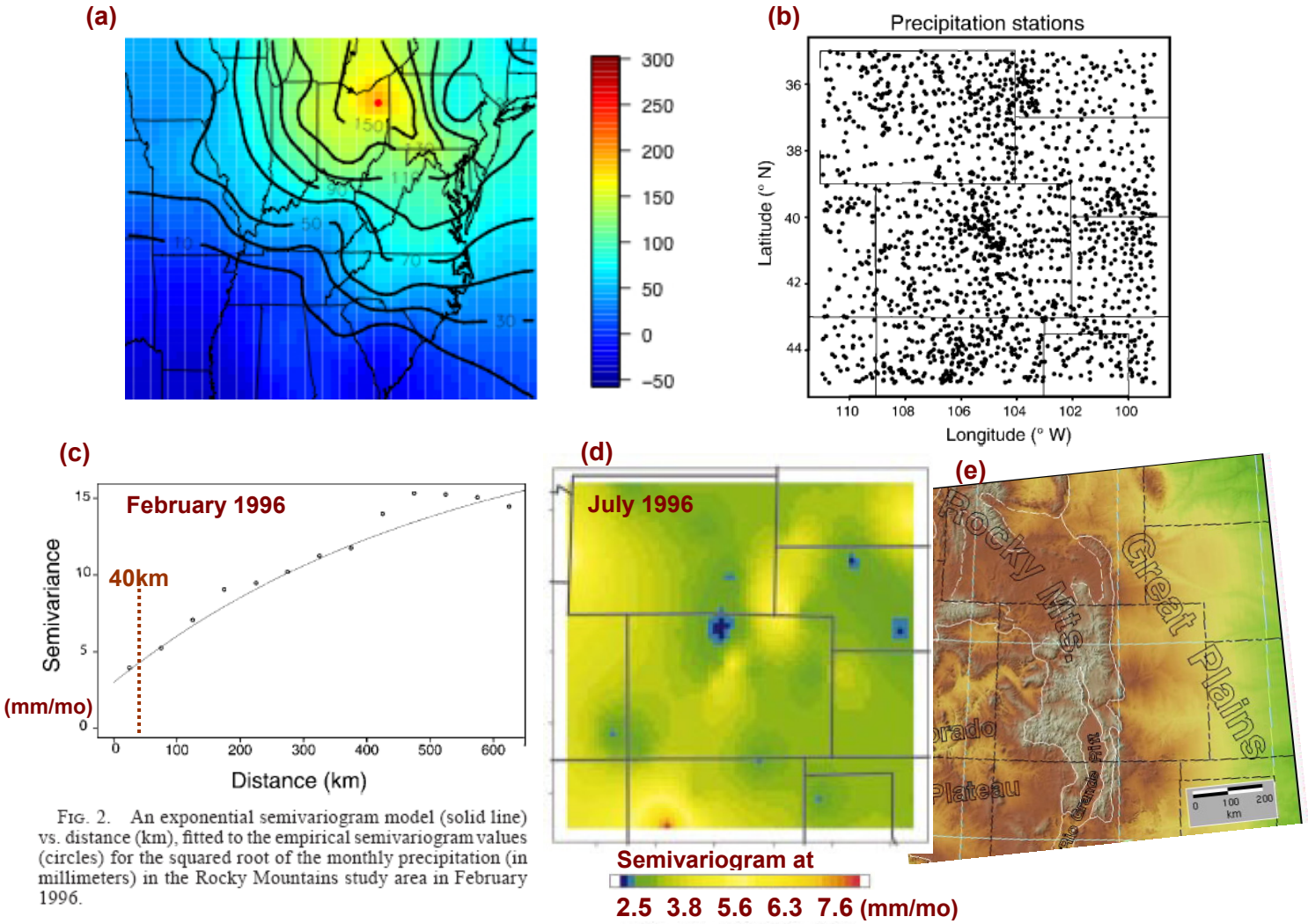

FIG. 2. An exponential semivariogram model (solid line) vs. distance (km), fitted to the empirical semivariogram values (circles) for the squared root of the monthly precipitation (in millimeters) in the Rocky Mountains study area in February 1996.

Figure 21. Spatial analyses showing regional connections. (a) Point-covariance map for surface ozone of the eastern U.S. between the grid point indicated by the red dot (in northeast Ohio) and the rest of the domain (from Matsuo 2005). Covariance units are [(ppb O_3)^2]. (b-e) Spatial autocorrelation model of monthly precipitation across the Central and Southern Rocky Mountains (Colorado is the state in the center of all maps). (b) Stations used in the analysis. (c) Semivariogram (mm/mo)[165] vs. spatial separation (km) at a given time, February 1996 – showing increasing differences (spatial variance) in precipitation between stations as inter-station distance increases. (d) Map of July 1996 precipitation semivariograms across the region at a set distance, 40km [marked with a dotted line in (c)]. Areas with lower semivariogram values have stronger spatial connectivity. (e) Topographic relief[166] for the domain in (b, c, d). (b-d from Fuentes et al. 2006; brown annotations added)

Discussion. Point-covariance map (a) shows the spatial extent of sites that have a strong correlation in time with the focus site's ozone dynamics. The text (§4.7.1) discusses this type of presentation in terms of point-correlation maps, which impart the same information – recall that covariances and correlations are directly related.[167] (c) and (d) show how monthly precipitation at any place in the domain is related to that of the surrounding region: in (c) as a function of distance between stations – spatial variance increases as spacing increases, and in (d) as a function of location within the domain – spatial variance is generally higher in regions with highest topographic heterogeneity (in e). We also see in (d) that the semivariogram at 40km is anisotropic, i.e., that spatial correlation depends on compass direction. (Link to text §4.7.2)

---

[165] Note re units: Units for semivariance are the square of the variable's units. In this example, however, semivariance units are 'mm' rather than 'square mm' because precipitation was first square-root transformed to better meet analysis requirements.

[166] Image from Jones et al., 1996, *Nature* 381:37-41. http://cires.colorado.edu/people/jones.craig/GSA/slide11_big.JPG

[167] Correlation(x,y) = covariance(x,y)/[SD(x)*SD(y)]

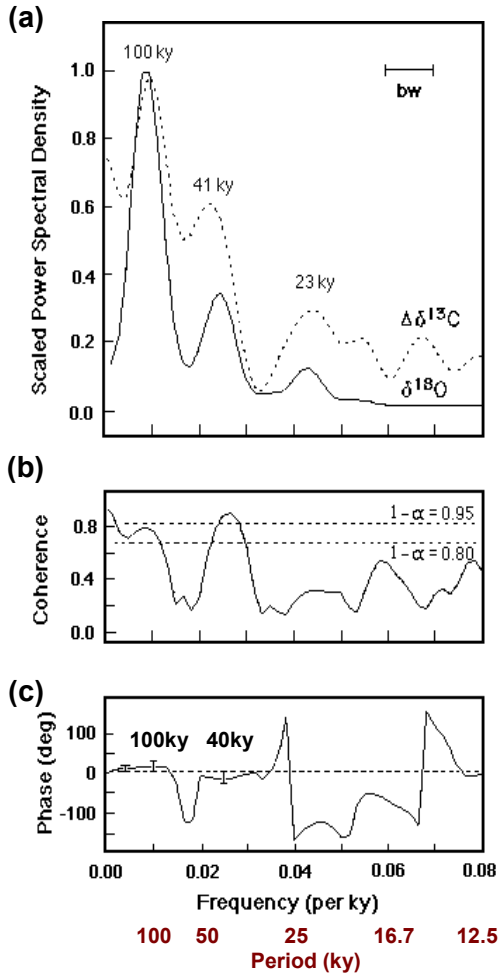

Figure 22. Cross-spectral Fourier analysis of two oceanic regional paleoclimate proxies, ^{18}O and ^{13}C isotopic variables. (a) Power spectra for the two individual timeseries, (b) their spectral coherence (h^2), and (c) phase.[168] Significance levels (5 and 20%) for coherence peaks are shown in (b). In (c), confidence intervals (C.I.; vertical 'I' bars) for the phase of major coherence peaks in (b) test if the phase is significantly different from 0º. Phase is presented in terms of degrees, with cycles defined as in phase at 0º and of opposite phase at 180º. In *x*-axis frequency and period units, 'ky'=10^3 years. Analysis bandwidth is shown in the upper right ('bw'=0.01/ky).

Discussion: The two climate proxy records share significant coherent (b) and in-phase (c) spectral power at periods of ~40 and ~100ky. Coherence values for these peaks ($h^2 \geq 0.80$) indicate that 80% or more of variance in these frequencies is shared (§4.7.3). In c, note that their C.I. bars extend to 0º. The periods of these peaks in ocean regional climate proxies suggest climate links with Earth orbital dynamics as they closely correspond to Milankovitch orbital parameter cycles for tilt (41ky) and eccentricity (100ky; Imbrie et al. 1992, 1993). (Link to text §4.7.3)

[168] Image from: McDuff, R.E., and G.R. Heath. 2001. Phase Relationships of Proxy Variables. Oceanography 540. Marine Geological Processes. Course web page. University of Washington, School of Oceanography. http://www2.ocean.washington.edu/oc540/lec01-26/ (brown *x*-axis labels added)

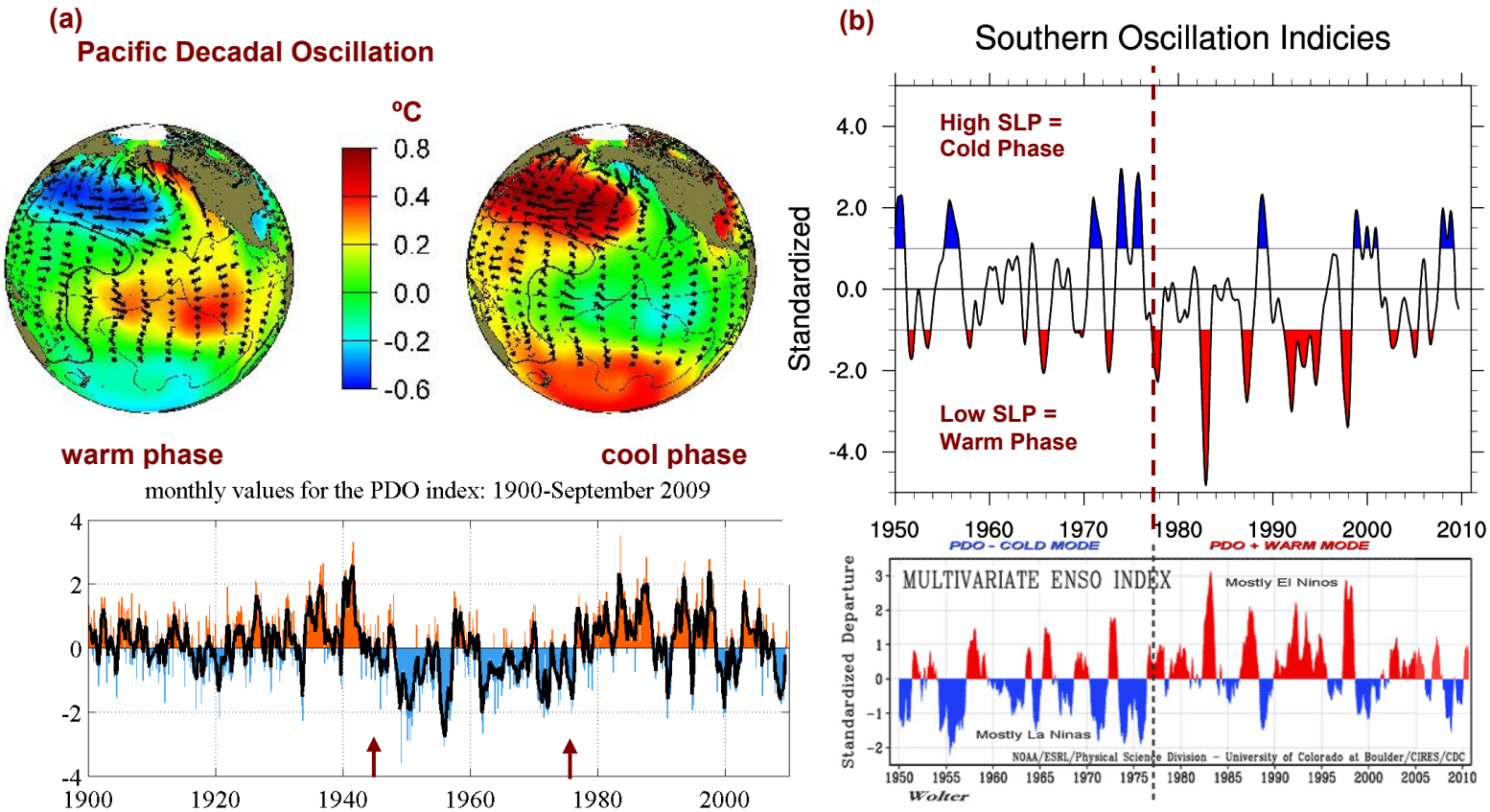

Figure 23. Hemispheric climate oscillations' spatial and temporal signatures: (a) The Pacific Decadal Oscillation (PDO) characteristic warm (*top left panel*) and cool (*right*) phase conditions across the Pacific: wintertime anomaly patterns in sea surface temperature (SST; color zones, ºC), sea level pressure (SLP, contours) and surface windstress (arrows). (*Bottom*) PDO Index monthly timeseries (standardized SST anomalies for the first principal component, §4.8.2).[169] (b) El Niño-Southern Oscillation (ENSO) timeseries since 1950 for two indices: Southern Oscillation Index (SOI; based on SLP anomalies) and Multivariate ENSO Index (MEI; based on SLP, SST, surface wind, surface air temperature, and cloudiness anomalies).[170] Plots are presented with the *x*-axes' year scales lined up. The plots in (b) illustrate two different approaches to data blocking: *(top panel)* High magnitude SOI events are selected using a threshold, *(bottom)* MEI is blocked by PDO cold vs. warm regime, which shifted after 1976 (vertical dashed line). For timeseries in both (a) and (b), blue = cool phase, red/orange = warm phase [note in (b) that SOI phases are of opposite sign than El Niño/La Niña (warm/cold) temperature anomalies]. (In a-b, brown annotations added)

Discussion: PDO phases are termed 'warm' and 'cool' (or 'cold') based on SST anomalies in central tropical to northeast extratropical Pacific (off of North American west coast) (a, *top*). Wind anomalies reverse directions in phase with temperature shifts. The PDO Index timeseries (a, *bottom*) shows the multidecadal persistence of these phases, with regime shifts following 1946 and 1976 (arrows along *x*-axis; see Figure 16b). In (b), note that the two ENSO indices capture the main events, though their relative structures vary. Both indices show the prevalence of cold ENSO events during the PDO's cold regime and warm ENSO in the PDO's warm regime – a pattern that suggests constructive interference between the two dynamics (Biondi et al. 2001). (Link to text §4.8.1)

[169] Images from: http://jisao.washington.edu/pdo/ (timeseries through September 2009, periodically updated)
[170] SOI image from: http://www.cgd.ucar.edu/cas/catalog/climind/soi.html (Trenberth 1984). MEI image from http://www.intellicast.com/Community/Content.aspx?ref=rss&a=126 (Wolter and Timlin 1993) through November 2006 – to this image is appended updated series through October 2009 from http://www.esrl.noaa.gov/psd/people/klaus.wolter/MEI/mei.html.

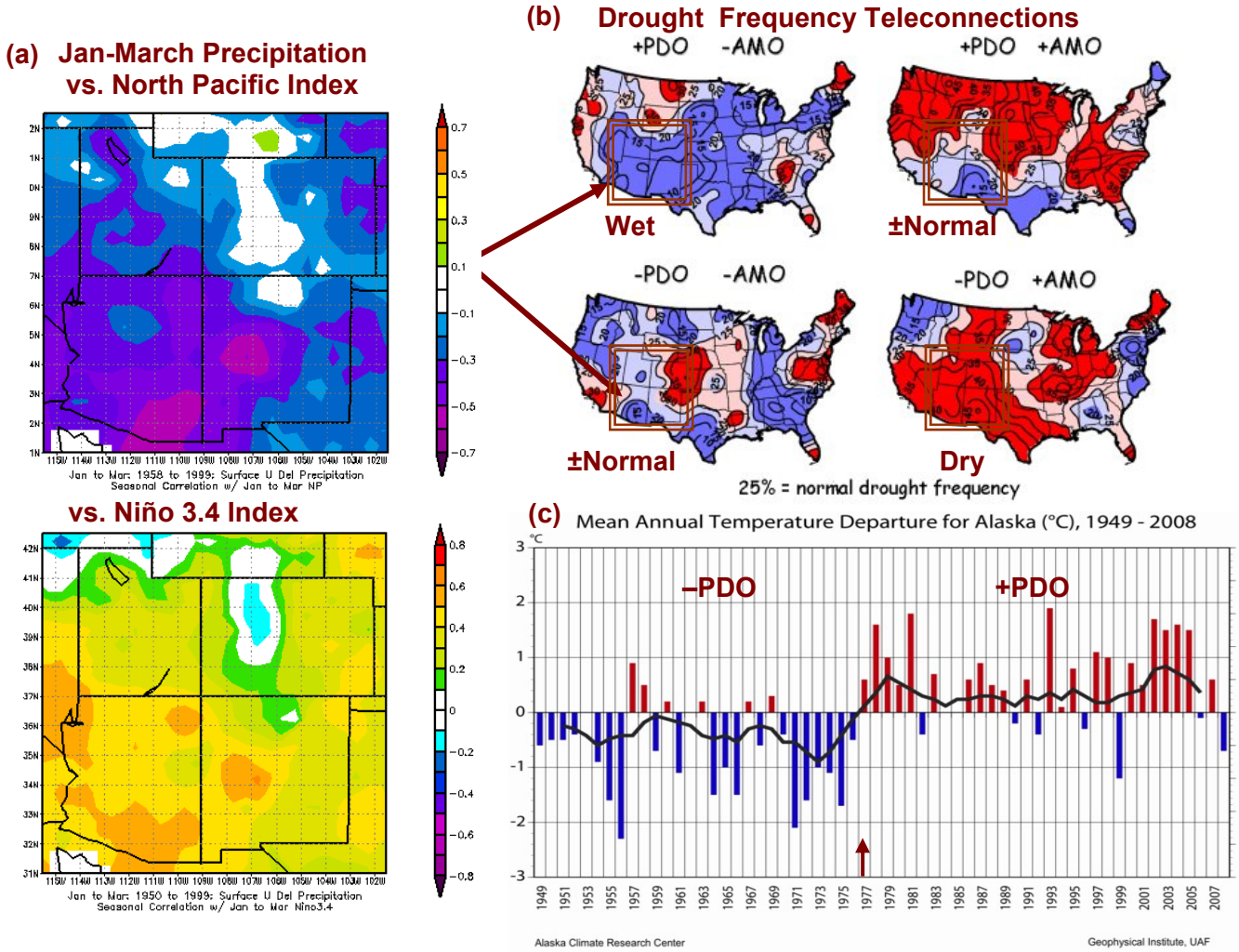

Figure 24. Pacific Ocean Basin teleconnections to conterminous U.S. and Alaskan climates. (a) Teleconnection correlation maps for January-March precipitation for the U.S. Four Corners region with *(top)* the North Pacific Index (NPI) and *(bottom)* Niño 3.4 Index (for 1958-1999 and 1950-1999, respectively). Warm colors = positive correlation, cold colors = negative. Plots generated on-line: http://www.esrl.noaa.gov/psd/data/correlation/. (b) Teleconnections for conterminous U.S. drought frequency as a function of the interaction between phases (+/–) of the Pacific Decadal Oscillation (PDO) and Atlantic Multidecadal Oscillation (AMO): blue areas = fewer droughts than normal, red = more (McCabe et al. 2004; image from Steward 2005, brown annotations added). (c) Mean annual temperature departure for Alaska (1949-2008) showing a dramatic shift in the late 1970's corresponding to the 1976/77 PDO regime shift (Figure 23a).[171]

Discussion: The negative correlation field in (a, *top*) indicates that a *negative* NPI in January-March corresponds to *high* precipitation into the American Southwest and Colorado River Basin. The underlying mechanism is that a negative NPI means a strong, deep Aleutian Low – this forces the mid-latitude jet stream and embedded winter storms to track farther south coming into North America, bringing moisture directly to the Southwest. In (b), these dynamics are similarly reflected in PDO teleconnections – *negative* NPI generally correspond to *warm* phases of the PDO ('+PDO') and so giving *wetter* conditions in the Southwest (b, *top*) than under –PDO (b, *bottom*). In (a, *bottom*), the *warm* SST phase of ENSO (positive Niño 3.4 Index) is *positively* related to Southwest precipitation in January-March. Warm SST's in the eastern Pacific are linked to greater rainfall in the eastern tropical Pacific and a weakened Subtropical High off the coast of northern Mexico – these lead to more precipitation in northern Mexico and the American Southwest. In (b), the +AMO phase brings drought across far more of the conterminous U.S. than under –AMO (*right* vs. *left panels*). However, which regions are affected by this drought is strongly determined by PDO phase. During the –PDO phase, when the Aleutian Low is weak, the mid-latitude jet stream is more zonal (following latitude lines) and brings moisture directly into the Pacific Northwest – keeping the +AMO drought out of this region (*top* vs. *bottom right panels*). This interaction could be evaluated statistically for a given region using multifactor techniques (§4.8.2). (Links to text §4.8, Table 3, Table 4)

[171] From: http://climate.gi.alaska.edu/ClimTrends/Change/TempChange.html (brown annotations added).