Dear Referees, Editor,

Thank you for the positive assessments of our paper ‘How over 100 years of climate variability may affect estimates of potential evaporation’ (Paper hessd-11-10787-2014). We appreciate the recognition of the relevance of the work and the thoroughness of the analysis.

Herewith, we provide a revised version of our manuscript in which we incorporated all suggestions and remarks made by you. To our opinion, incorporating all valuable suggestions has clarified and strengthened the manuscript, especially because we:

- provide more insight in the errors made in relation to the structure of the evaporation models
- put more emphasis on the quantification of the error that could be made by using crop factors
- give suggestions how to correct for such errors.

Detailed responses to all comments of the referees can be found in the following pages. The revised manuscript, with the changes highlighted, is provided at the end of this letter.

Yours sincerely,

Ruud Bartholomeus

Detailed replies to the referees' comments

Referee #1

#1_1: General Comments: The authors have done a lot of work characterizing the variability in ET estimates using various calibration periods over the last century. Their overall conclusion is that using calibrated coefficients extrapolated from a short period of time under different climate conditions can lead to systematic differences between empirical and process-based models. This is not a surprising result, and I would be interested to see the results presented in a way that gives readers tangible information that allows them to make the best decision of how to model ET given limited radiation or ET measurements.

Reply 1: We agree that it is widely acknowledged that the application of empirical coefficients is limited to their period of calibration. However, although this limitation of the two-step approach is known among both scientists and practitioners, the approach is still regularly applied in hydrological modeling studies on different spatial and temporal scales without appropriate consideration or warnings. Our study is novel in the sense that we provide quantitative information on the limitations of the two-step approach for different vegetation types and estimation procedures (potential evaporation) using a very long time series, allowing multiple 30-years periods to be assessed. Such an approach can be used in similar modeling studies to i) derive uncertainty ranges for the parameters, ii) quantify the errors that are introduced by a specific method and set of parameters, and iii) correct for the errors when they are predictable.

In the revised version of the manuscript we put more emphasis on the quantification of the potential error and how one could correct for this error, rather than simply stating that applying empirical coefficients for extrapolations may introduce errors. We added a section in which we provide more guidelines on the choices one should make in evaporation modeling. See p.1., l. 10-11 and l. 21-24; p. 6-7, l. 31-12; p. 18, l. 8-29; p. 21, l. 17-23.

#1_2: I think the paper would benefit from an additional section examining the reliability of published crop coefficients and commonly used parameters for ET estimations over the period of record, and draw some general conclusions about that. At the very least, the authors should include more context for the estimated parameters generated in this study in terms of how they compare to already published values (they cite Feddes, 1987 and Allen, 1998 – others to look at could include Shuttleworth, 1992 or other ET factors in hydrology reference texts).

Reply 2: We agree that comparing crop factors that are being used for the meteorological conditions in the research area, i.e. the Netherlands, could be a valuable addition to our analysis. Therefore we added a comparison of crop factors to the revised manuscript (see p. 17, l. 6-19). It should be realized, however, that comparison with published crop factors is misleading, as these are obtained for the non-calibrated Makkink reference evaporation. Nevertheless, besides comparing model derived crop factors with measured ones, we show the variability in crop factors, caused by changing climatic conditions. A comparable variability can be expected for published crop factors. The analysis thus provides insight into the uncertainty ranges that can be expected for published empirical coefficients and this information can be used to better judge the uncertainty in the results of a modeling exercise (see p. 17, l. 19-23).

#1_3: In addition, a number of the figures are difficult to read – I'd suggest presenting a representative figure or few figures from some of the multiple-pane plots and explaining the differences between groups in the text.

Reply 3: We have already limited the number of panes by not showing all results, for all crops and all evaporation components. In our opinion, further limiting the number of panes might hamper the clarity of the results. We therefore decided not to remove figures from the multiple-pane plots.

#1_4: Specific comments: Figure 2: What is the significance of the dashed lines compared to the solid lines?

Reply 4: These lines were only dashed to clarify the specific connection; we clarified the connection in the revised manuscript. See Figure 2.

Referee #2

General comments

#2_1: This paper investigates how non-stationarity in climate data can influence the estimates of potential Evaporation using the “two step” or crop factor approach. Overall this is a timely discussion to have. It is more and more clear that there is a large amount of variation in the climate and this affects the performance and behavior of hydrological and climate modelling if parameters in the model are considered stationary. Simply put, non-stationarity is unexplained variance. On the one hand, it is good to indicate these issues and to warn practitioners, but on the other hand, do we really believe we can make accurate predictions outside a calibration period? This is not even true for simple regression models, so why would it be true for calibrated hydrological and climate models. Any extrapolation outside calibration data is going to suffer from increased uncertainty. This has been known for years. The question might be more, why is this easily forgotten, and how do we deal with it? The probable reason why it is easily forgotten, is that we believe that our models, because we are attempting to represent real physical processes, are not regression models.

Reply 5: Thank you for your comments, which are related to the comments made by reviewer #1 (see Reply 1). We fully agree that although experts know that the two-step approach introduces errors, many seem to not be aware of which methods they are actually using and what the consequences may be for their modeling studies. The limitations of the two-step approach are often neglected, and apparently there is a need to demonstrate its limitations. We included additional text on this topic in the revised manuscript.

We agree with the reviewer, that applying coefficients outside their calibration range is a major flaw in research. We believe a primary reason for this recurring problem is that warnings about extrapolation are often qualitative and therefore extrapolation sensitivity can be disregarded as “noise” within a larger “signal”. This shows the importance of this research – by quantifying the sensitivity of evaporation to this commonly overlooked assumption, we hope to stop the propagation of this error in future studies.

With our analysis we now quantify the potential errors, which provides insight in the reliability of the method. Such analysis supports both scientists and practitioners to decide which method is appropriate for their analysis. See Reply 1 for the corrections that are made in the revised manuscript.

#2_2: What I really missed in the paper is a solution. We could define the uncertainty and attempt to adjust the management to deal with the uncertainty, but this is rather unsatisfactory as a scientist. The other, more important approach, is to find a way to modify the model to deal with the issue. Are you suggesting we throw out the two-step approach? Or can we adjust the two-step approach? In the end, Figure 10 actually indicates that there is some pattern in the over and underestimation, both between models and in time periods. So there is some predictability in the actual deviations. This would have been nice to explore.

Reply 6: Thank you for this valuable suggestion. We extended our analysis to provide guidelines to estimate the potential error that is made in extrapolations, based on the differences in climatic conditions between reference period and application period. See p. 18, l. 8-19.

#2_3: The other issue of interest that emerges from the paper is the comparison between models. While this is highlighted (Hargreaves and Blaney-Criddle versus Makkink and Priestley-Taylor), it is not really analysed in relation to the structure of these models. Why do the temperature models fail more than the radiation driven models?

Reply 7: We explain the differences in the revised manuscript, supported by statistical correlations between parameters used in the different E_{ref} methods and E_{ref_PM} trends. See p. 12-13, l. 28-6.

#2_4: Finally there is the difference between vegetation. While this is just synthetic data, this incorporates the “current knowledge” about the evaporation from these vegetation types. In addition, the variation between veg types appears to be lower than between models. Is this interesting?

Reply 8: The last paragraph of section 3.2 explains differences between vegetation types, based on their structure. We would like to note that Figure 7 actually shows that the variation between vegetation types is larger than the variations between models.

#2_5: So, while I think the analysis is tidy and neat, and the topic of interest, I miss depth in the article to actually progress the science and the application.

Reply 9: We believe that the suggestions of the referees and the additional analysis included in the revised manuscript strengthened the scientific aspects of our analysis and now provides sufficient guidelines for both scientists and practitioners to quantify and minimize potential errors induced by simple regression models and empirical coefficients in the two step approach for estimating potential evaporation.

Specific comments

#2_6: I have a few specific comments P10792 line 27: no-analogue? Is this a typo, I wasn't sure, should this be non-analogue?

Reply 10: no-analogue is the correct term. See e.g.
http://www.sciencedirect.com/science/article/pii/S0921818112002299;
http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0006825#pone-0006825-g003

#2_7: P10795 line 5 & 6: The accuracy of SWAP, It is not really irrelevant. I think you need to at least identify whether the choices of parameters in SWAP would affect the variability and the relative proportions of the calculated E components. So has your choice of crop, soil depth etc affected the different E component variation in time. You are assuming that the relative relationship between Ei and other E components is invariant of your crop choice and soil depth. Page 10797 line 14, this might cover my previous comment, but still worth checking.

Reply 11: We selected vegetation types ranging from grasses, to shrubs and forests to demonstrate that our findings hold for different vegetation structures. Each of these vegetation types has its own specific parameter values. Therefore, these different vegetation types already include different choices of parameters in SWAP that affect the variability and relative proportions of the calculated E components. We clarified this in the revised manuscript (see p. 8, l. 21-24). Additionally, Ei is not invariant of crop choice, as the simulated interception is vegetation dependent. As already indicated by the referee, we took standard values for these vegetation classes as used for the National Hydrological Instrument for the Netherlands. Soil depth is not relevant, as we only consider potential evaporation.

#2_8: Page 10799 line 24: Would it worth highlighting what in these models causes this? They are both calibrated on the same data, both temperature based, but given the same temperature series one deviates downward (under climate change) and one upward, even though the temperature series has the same direction for both. Looking at the equations in Table 1, both use average temperature (which is supposedly increasing), but Har also uses Radiation and the difference between Tmax and Tmin, which might be stable

Reply 12: This is an interesting observation; the different directions in change for BC and Har are caused by a general decrease in Tmax-Tmin, while the mean temperature increases. This is added to the revised manuscript (see p. 13, l. 7-9).

#2_9: Page 10805 line 7: advance in the ability

Reply 13: This has been corrected to advance in the abilities (see p. 20, l.1).

#2_10: Page 10805 line 12: assumptions (plural)

Reply 14: Corrected (see p. 20, l. 6)

Referee #3

#3_1: This paper evaluates the sensitivity of the two step approach to calculate evaporation to the length of the calibration period and the chosen reference years. It compares four different two step evaporation methods with the Penman-Monteith method and compares these five methods with potential evaporation obtained with the process based SWAP model for four vegetation classes. The analysis shows that the empirical equations are highly sensitive to the length of the calibration period and the timing of the selected period and are therefore hard to transfer in time to use for example in climate impact assessments.

General comments: The paper is written very clearly, especially the introduction that provides a very good setting for the paper. To my opinion the description of methods and results misses some background information which I will further detail below. The lengthy dataset used is very valuable for this demonstration, yet this is also an ideal situation where all atmospheric variables are available. The authors could maybe elaborate a little more on what one could do when this information is not available, i.e. the Makkink and Priestley-Taylor methods seem to be doing relatively well.

Reply 15: Thank you for the positive response on the manuscript. In the revised manuscript we provide guidelines to predict the error that is made by using different methods, which have different data requirements (See Reply 1 and Reply 5 for the corrections that are made in the revised manuscript).

#3_2: Moreover, this paper only discusses a Dutch site, can this information be transferred to other locations on the globe or would the results be different for other climate zones?

Reply 16: The absolute values and differences are case specific and thus not applicable to other regions. Nevertheless, the sensitivities identified in this study are related to the models themselves and how they are affected by different climate fitting parameters. While projected changes in radiation and temperature vary globally, the general trends are consistent, and it is reasonable to expect that similar differences identified for this specific case can be expected for other climatic regions. We extended the description of the site and how it resembles global trends in the revised manuscript (see p. 9, l. 17-23).

The chosen site is unique in that it has a long enough historical record to allow for comparisons across different sub-periods. The majority of climate stations have much shorter records, which would not show the change in extrapolation errors through time.

#3_3: The discussion of SPEI values is very good, interesting to see the influence of the calculated evaporation on a relevant indicator. Overall the only drawback is that the results and conclusions are not really novel information.

Reply 17: This comment is similar to those raised by Reviewer #1 and #2 and addressed in Reply 1 and Reply 5).

Specific comments:

#3_4: - The paper provides figures and information of the newly calibrated two step approaches. It is unclear how the results compare to the un-calibrated equations with default values from literature. The same applies for the calibration of crop factors. How do these compare to crop factors from literature and how does the calculated evaporation compare to evaporation calculated using these standard values?

Reply 18: We now provide information on calibrated E_{ref} parameters and compare obtained crop factors with those from literature. See Reply 2 and p. 17, l. 6-23.

#3_5: - The variables involved in calibration are very briefly mentioned in section 2.3 for the reader it is hard to see to which equation these apply. Maybe also mark the variables bold in the equations in Table 1.

Reply 19: The calibration variables are also presented and explained in Table 1, which we now clarified by marking them bold. See Table 1.

#3_6: - In the introduction the authors mention a multiplication factor of 1.1 – 1.3 if interception is involved – has this factor been considered in the remainder of the study? Could the (non)-use of this factor influence the results?

Reply 20: This factor should only be used in combination with Kt and if interception is not simulated explicitly. Therefore, because we simulated interception explicitly, in our study the multiplication factor has not been used. We clarified this in the revised manuscript (see p. 5, l. 30).

#3_7: - Can the calibration or set-up of the SWAP model be considered stationary over time and does this influence the analysis?

Reply 21: Considering stationary vegetation does not affect the results, as we study potential evaporation in time for different vegetation classes instead of specific sites with a dynamic vegetation. Additionally, simulating dynamic vegetation and herewith succession, would unnecessary complicate the analyses.

#3_8: - Section 4 is structured in a non-logical order. I would suggest to either add section 4.2 and 4.3 to the results section or move 4.1 to the end of section 4.

Reply 22: We moved section 4.2 to the results sections, but kept section 4.3 in the discussion section, as 'implications' fit best there. See p. 15, l. 11-30

Corrections:

#3_9: - Both data sets and datasets are used

Reply 23: the occurrence of data set has been corrected to dataset (see p. 9, l. 29).

#3_10: - Section 3.1 Deviation deceases should read Deviation decreases

Reply 24: Corrected (see p. 13, l. 21).

Sensitivity of potential evaporation estimates to 100 years of climate variability ~~How over 100 years of climate variability may affect estimates of potential evaporation~~

R.P. Bartholomeus[1], J.H. Stagge[2], L.M. Tallaksen[2], J.P.M. Witte[1,3]

1 KWR Watercycle Research Institute, P.O. Box 1072, 3430 BB Nieuwegein, The Netherlands

2 Department of Geosciences, University of Oslo, P.O. box 1047, Blindern, 0316 Oslo, Norway

3 VU University, Institute of Ecological Science, Department of Systems Ecology, de Boelelaan 1085, 1081 HV Amsterdam, The Netherlands

Correspondence to: R.P. Bartholomeus (ruud.bartholomeus@kwrwater.nl)

Abstract

Hydrological modeling frameworks require an accurate representation of evaporation fluxes for appropriate quantification of e.g. the water balance, ~~the~~ soil moisture budget, ~~droughts,~~ recharge and groundwater processes. Many frameworks have used the concept of potential evaporation, often estimated for different vegetation classes by multiplying the evaporation from a reference surface ('reference evaporation') with crop specific scaling factors ('crop factors'). Though this two-step potential evaporation approach undoubtedly has practical advantages, the empirical nature of both reference evaporation methods and crop factors limits its usability in extrapolations ~~and~~ under non-stationary climatic conditions. In this paper, rather than simply warning about the dangers of extrapolation, we ~~assess actually~~ quantify the sensitivity of potential evaporation estimates for different vegetation classes using the two-step approach when calibrated using a non-stationary climate. We used the past century's time series of observed climate, containing non-stationary signals of multi-decadal atmospheric oscillations, global warming, and global dimming/brightening, to evaluate the sensitivity of potential evaporation estimates to the choice and length of the calibration period. We show that using empirical coefficients outside their calibration range may lead to systematic differences between process-based and empirical reference evaporation methods, and systematic errors in estimated potential evaporation components. ~~Such extrapolations of time-variant model parameters are not only relevant for the calculation of potential evaporation, but also for hydrological modeling in general, and they may limit the temporal robustness of hydrological models.~~ Quantification of errors provides a possibility to correct potential evaporation calculations and to rate them for their suitability to model climate conditions that differ significantly from the historical record, so-called no-analogue climate conditions.

1 Introduction

Evaporation from the vegetated surface is the largest loss term in many, if not the most, water balance studies on earth. As a consequence, an accurate representation of evaporation fluxes is required for appropriate quantification of surface runoff, the soil moisture budget, transpiration, recharge and groundwater processes (Savenije, 2004). However, despite being a key component of the water balance, evaporation figures are usually associated with large uncertainties, as this term is difficult to measure (Allen et al., 2011) or estimate by modeling (Wallace, 1995).

Research attempting to model the evaporation process has a long history (Shuttleworth, 2007). This research took two parallel tracks, with the meteorological community developing process-based models of surface energy exchange and the hydrological community considering evaporation as a loss term in the catchment water balance (Shuttleworth, 2007). To quantify the evaporation loss term, many hydrological modeling frameworks have used the concept of potential evaporation (Federer et al., 1996; Kay et al., 2013; Zhou et al., 2006), defined as the maximum rate of evaporation from a natural surface where water is not a limiting factor (Shuttleworth, 2007). With the progression from catchment-scale lumped models (such as HBV (Bergström and Forsman, 1973)) to distributed models with increasing spatial resolution and spatially resolved data (such as SHE (Abbott et al., 1986)), the explicit representation of land surface water budgets also increased (Ehret et al., 2014; Federer et al., 1996). To this end, estimation of evaporation from a variety of land surfaces within the simulated domain is needed (Federer et al., 1996). More models were developed that included vegetation explicitly, commonly by describing the stomatal conductance of the vegetation as a function of environmental drivers (see Shuttleworth (2007) and references therein). However, until now these models are rarely used in practice and merely have a scientific meaning.

Parallel to this development, the irrigation engineering community refined the traditional potential evaporation approach (Shuttleworth, 2007). They developed the ‘two-step approach’ (Doorenbos and Pruitt, 1977; Penman, 1948; Zhou et al., 2006; Feddes and Lenselink, 1994; Vázquez and Feyen, 2003; Hupet and Vanclooster, 2001), in which the potential evaporation of a specific crop or vegetation class is estimated by multiplying the evaporation from a reference surface with empirical crop specific scaling factors: ‘crop factors’. This development was mainly driven by the need for a relatively simple approach using commonly available data from climate stations. The two-step approach has even expanded outside the field of irrigation engineering into hydrological modeling frameworks. Crop factors are now being applied in 1D hydrological models (e.g. Tiktak and Bouten (1994)), spatially lumped models (e.g. Driessen et al. (2010); Calder (2003)), and spatially distributed hydrological models (e.g. Ward et al. (2008); Shabalova et al. (2003); Trambauer et al. (2014); Van Roosmalen et al. (2009); Lenderink et al. (2007); Bradford et al. (1999); Guerschman et al. (2009); Sperna Weiland et al. (2012); Van Walsum and Supit (2012); Vázquez and Feyen (2003)).

With the development of the two-step potential evaporation approach, different equations to simulate reference evaporation have been suggested (Federer et al., 1996; Bormann, 2011; Shuttleworth, 2007) for use in both regional and global hydrological models (e.g. Sperna Weiland et al. (2012); Haddeland et al. (2011)). However, due to their empirical nature, these equations are limited in their transferability in both time and space (Feddes and Lenselink, 1994; Wallace, 1995). Since the increasing need for predictions under global change (land use and climate) (Ehret et al., 2014; Coron et al., 2014; Montanari et al., 2013), the empirical nature of most commonly used potential evaporation approaches is a serious drawback (Hurkmans et al., 2009; Wallace, 1995; Shuttleworth, 2007; Witte et al., 2012). Thus, although the two-step approach may be warranted for practical reasons, both the reference evaporation and estimated crop factors include a series of empirical parameters that may affect the validity and general applicability of the estimated potential evaporation for a specific vegetation class.

Since the term “potential evaporation” has been used by the hydrologic community to refer to several different combinations of evaporation components in the past, it is important to re-introduce these definitions and to be very specific about nomenclature in future evaporation research. Total evaporation (E_{tot}) from a vegetated surface is the sum of three fluxes: transpiration (E_{t}), soil evaporation (E_{s}) and evaporation of intercepted water (E_{i}). E_{t} and E_{s} occur at a potential rate when the availability of water (soil moisture or interception) is not limiting. As we will only focus on potential rates in this paper, all values should be interpreted as potential, unless stated otherwise. Reference evaporation (E_{ref}) is defined as the rate of evaporation from an extensive surface of green grass, with a uniform height of 0.12 m, a surface resistance of 70 s m^{-1}, an albedo of 0.23, actively growing, completely shading the ground and with adequate water (Allen et al., 1998). By definition, E_{i} is not part of reference evaporation, as it is defined for a plant surface which is externally dry (Federer et al., 1996;

Allen et al., 1998). Often, the term reference evapotranspiration is used instead, which is the sum of transpiration (E_t) and soil evaporation (E_s). By definition (Allen et al., 1998) the reference crop completely shades the ground and hence E_s will be zero and E_{ref} equals E_t of the reference crop (at least for daily estimates, when the soil heat flux can be assumed zero). This is in agreement with the definition of Penman (1956) who also stated that the often-used expansion of the term "reference evaporation" to "evapotranspiration" was unnecessary.

E_{ref} is used in the two-step method to estimate the potential evaporation, E_p, of a crop or vegetation stand. E_p will reduce to the actual evaporation, E_a, in case of water shortage or waterlogging. Here, we focus on the estimation of E_p from E_{ref}, by multiplying E_{ref} with a crop factor K (Allen et al., 2005; Feddes, 1987; Allen et al., 1998; Penman, 1956). Different applications of crop factors exist:

- K_t corrects for potential transpiration of a crop with a dry canopy only, i.e. $E_t = K_t \times E_{ref}$. This corresponds to the basal crop factors defined by Allen (2000), which are equivalent to the approach of Penman (1956).
- K_{ts} corrects for both potential transpiration and potential soil evaporation for a crop with a dry canopy, i.e. $E_t + E_s = K_{ts} \times E_{ref}$. This corresponds to the single crop factors defined by Allen et al. (1998).
- K_{tot} corrects for potential total evaporation, i.e. transpiration, soil evaporation, and interception. Using K_{tot} with E_{ref} directly gives E_{tot}, i.e. $E_{tot} = K_{tot} \times E_{ref}$.

K_{tot} holds for crop factors that have been derived by soil water balance experiments, and especially from sprinkling experiments in the field, where water is applied in such quantities that soil water is not limiting for plant growth (Feddes, 1987). Sprinkling, however, leads to interception. So, crop factors like those of Feddes (1987) implicitly involve E_i. Therefore Feddes (1987) emphasizes that the presented crop factors "are averages taken over a population of 'average', 'dry', and 'wet' years, that will certainly not be homogeneously distributed". The crop factor approach by Feddes (1987) is different from the single crop factor approach of Allen et al. (1998), as crop factors from the latter are by definition applied to correct for $E_t + E_s$, or for E_t only (Allen, 2000). However, Allen et al. (1998) indicate that their crop factors should be multiplied with a factor 1.1-1.3 if interception, due to sprinkling irrigation for example, is involved (i.e. if interception is not simulated explicitly). This indicates that E_i could significantly affect potential evaporation from a vegetated surface. As E_i is largely driven by precipitation, a term that is generally not incorporated in E_{ref} methods, it has already been stated that the crop factor approach only makes sense in times of drought, when interception does not contribute to the total evaporation (De Bruin and Lablans, 1998). This condition is especially relevant for tall forests, which intercept a higher percentage of rain water, under climatological conditions with significant rainfall (De Bruin and Lablans, 1998). Nevertheless, this crop factor approach is used in practice (Van Roosmalen et al., 2009).

The objective of this paper is to assess the sensitivity of potential evaporation estimates for different vegetation classes using the commonly used two-step approach when calibrated based on a non-stationary climate. To this end, we use century long meteorological observations representing the historic variability in climatic conditions at the De Bilt, The Netherlands climate monitoring station. The past century's global warming, dimming and brightening periods (Suo et al., 2013; Stanhill, 2007; Wild, 2009; Wild et al., 2005), and their effects on evaporation provide an opportunity to evaluate the robustness of the two-step estimation of potential evaporation for non-stationary conditions. Given the 20th century climate induced variability in E_{ref} and the projected increase for the near future, which has no historical analogue,~~no analogue ongoing increase for the near future~~ (Fig. 1), it is of great importance to recognize the limitations of applying empirical coefficients outside their calibration range (i.e. extrapolation). This applies not only to transferring coefficients in space, as between climatic regions (Allen et al., 1998), but also in time.

The 20th century global surface temperature can be characterized by two major warming periods; the first one from about 1925-1945, followed by a period of cooling, and a second starting in about 1975 and continuing to the present (Jones and Moberg, 2003; Yamanouchi, 2011). While the variations in temperature until the 1970s can be related to changes in global radiation, i.e. global dimming and brightening, this relationship no longer holds for the rapid warming since 1975 (Wang and Dickinson, 2013). Empirical equations for reference evaporation that use either radiation or temperature implicitly assume a relationship between the two variables. Given the nonlinearity of evaporation components, it is not only questionable whether empirical equations for reference evaporation will be applicable under future climatic conditions (Shaw and Riha, 2011), but also whether they are applicable for the recent past.

Although the limitations of using empirical coefficients to calculate evaporation are generally well known, the potential errors that could be made by using such coefficients in evaporation calculations have, as far as we know, never been quantified. Thus, there is a need to raise the awareness of the uncertainty that may result applying such an empirical estimation method outside its valid area (site and time specific). In this study we systematically unravel the use of the two-step approach to simulate potential evaporation and identify and actually quantify systematic errors that may be introduced when empirical coefficients are applied outside their calibration period. Such extrapolations of time-variant model parameters are not only relevant for the calculation of potential evaporation, but also for hydrological modeling in general, thus limiting the temporal robustness of hydrological models (Ehret et al., 2014; Karlsson et al., 2014; Coron et al., 2014; Seibert, 2003). Quantification of errors, as demonstrated in this study, provides the possibility to i) derive uncertainty ranges for the parameters, ii) quantify the errors that are introduced by a specific method and set of parameters, and iii) correct for the errors when they are predictable.

2 Methods

2.1 General approach

We use 108 years of meteorological observations to quantify the sensitivity of potential evaporation when calibrated using a non-stationary climate for various natural vegetation classes using the two-step approach. We investigate how empirical E_{ref}-methods and empirical K-values affect the validity of the estimated potential evaporation for different vegetation classes, by applying empirical coefficients outside their calibration period. We vary the calibration period in both length (2-30 years) and reference period (in 1906-2013).

First (section 2.3), we simulate reference evaporation according to the process-based Penman-Monteith equation ($E_{\mathrm{ref\_PM}}$), which is considered the international standard method for estimating reference evaporation (Allen et al., 1998). In addition, we apply four empirical equations that contain constants derived for a calibration period (Fig. 2: §2.3). From these simulations, we identify deviations between each empirical E_{ref} method and the $E_{\mathrm{ref\_PM}}$ (Fig. 2: §2.3).

Secondly (section 2.4), we generate time series of the main components of potential evaporation, i.e. synthetic series of E_t, E_s and E_i, for five different vegetation classes, using the Soil-Vegetation-Atmosphere Transfer (SVAT) scheme SWAP (Kroes et al., 2009; Van Dam et al., 2008) (Fig. 2: §2.4). SWAP allows users to simulate potential evaporation for different vegetation classes directly (i.e. one-step approach), by parameterizing the Penman-Monteith equation for each vegetation class implicitly rather than using crop factors. These synthetic series are considered 'observations' throughout the paper for all comparisons with estimates from the two-step approach.

Finally (section 2.5), we derive monthly crop factors for each vegetation type (5x) and for each E_{ref} method (5x) based on the synthetic data of E_{t}, E_{s} and E_{i} for a calibration period (e.g. 1906-1935) to simulate crop factor estimation using field measurements (Fig. 2: §2.5). We use different (3x) definitions of crop factors: for transpiration (K_{t}), for transpiration plus soil evaporation (K_{ts}) and for total evaporation (K_{tot}). Next, we apply the two-step approach, using E_{ref} and crop factors from the calibration period to calculate daily 'predicted' evaporation components (3x) for each vegetation class (5x) and each E_{ref} method (5x) for the entire period (1906-2013) (Fig. 2: §2.6). Doing so, the empirical E_{ref} methods and crop factors are applied outside their calibration range. From these simulations we quantify the deviations introduced by the use of E_{ref} and K, by comparing the evaporation components obtained with the two-step approach to the synthetic 'observations' (Fig. 2: §2.6). Each of these steps, which are executed for all calibration periods during the period 1906-2013 (2697x), are described in greater detail in subsequent sections.

Although SWAP may be expected to provide adequate evaporation values, its absolute accuracy is not discussed in this paper, because we focus on the sensitivity of the two-step approach using synthetic (hypothetical) data only. Therefore, the actual accuracy of SWAP is irrelevant for this paper. To ensure that our analysis is not biased by a specific choice of SWAP parameter settings, we considered different vegetation classes ranging from grasses to shrubs and forests. Doing so, we include different parameter sets that affect the variability and relative proportions of the calculated E components. For a detailed discussion of the SWAP model and its accuracy, please refer to Kroes et al. (2009) and Van Dam et al. (2008). By comparing potential evaporation components obtained from the two-step approach with the synthetic 'observations' as simulated using the physical SWAP model, we are able to quantify the deviations introduced by using different E_{ref} methods in combination with crop factors, as no other source of uncertainty is involved.

2.2 Meteorological data

We use meteorological data from De Bilt, The Netherlands, covering the period 1906-2013, which was provided by the Royal Netherlands Meteorological Institute (KNMI). De Bilt (longitude = 5.177° east, latitude = 52.101° north, altitude = 2 m) is the main meteorological site of the KNMI, located in the center of the Netherlands. Daily records are available for minimum and maximum temperature, sunshine hours, wind speed, and precipitation from 1906 onwards, and for global radiation from 1957. The observations are continuous, except for April 1945, where values from April 1944 are used instead. All required input variables are calculated for the period 1906-2013 following Allen et al. (1998). Observed global radiation was used to derive the Angstrom coefficients needed to calculate daily global radiation (Allen et al., 1998) from 1906 onwards. For consistency we only use these simulated values for further analysis, which agree very well with observations (1957-2013, $R^2_{adj} = 0.96$). Wind speed, measured at different heights, was scaled to the reference height of 2 meter following Allen et al. (1998) and corrected for systematic differences between measurement periods. Fig. 3 shows the annual values and the 30 year moving averages of the variables used to calculate evaporation from De Bilt.

Although the results are only valid for the site and period they were developed, the times series of radiation for De Bilt station resembles the global trends of global dimming/brightening. Values of global radiation (R_s) from De Bilt show a similar trend to the observations for Stockholm, as presented in Wild (2009). The data (Fig. 3) show an increase in temperature consistent with previous studies (Solomon et al., 2007) and a pattern of sunshine duration consistent with dimming and brightening for northwestern Europe identified by Sanchez-Lorenzo et al. (2008).

Long time series of meteorological observations will, to some extent, not be homogeneous, for example due to changes in measurement devices over time. However, this does not affect the calculations herein, as the aim is to investigate the sensitivity of the two-step potential evaporation methodology to non-stationary climate, rather than to produce an exact reconstruction of the last century's climate conditions. In this way, changes in measurement accuracy with time simply represent another non-stationary trend in this data-set.

2.3 Reference evaporation

Several methods are available for calculating reference evaporation, differing in complexity and empiricism (Sperna Weiland et al., 2012; Bormann, 2011; Federer et al., 1996). Here we analyze five of these methods, given in Table 1: the physically-based Penman-Monteith equation (PM), the radiation based methods of Makkink (Mak) and Priestley-Taylor (PT), and the temperature based methods of Hargreaves (Har) and Blaney-Criddle (BC).

The FAO-56 method (Allen et al., 1998), using PM parameterized for reference grass, is recommended as the international standard for calculation of E_{ref}. Given the physical basis of PM, it can be used globally, without the need to estimate or calibrate its parameters (Droogers and Allen, 2002). In contrast, the methods of Mak, PT, Har, and BC contain empirical coefficients, derived for specific meteorological conditions and sites. Following Farmer et al. (2011) we consider $E_{\mathrm{ref\_PM}}$ as the best approximation of E_{ref}. In order to reduce any systematic differences between E_{ref} values, we estimate the empirical factors C_1, C_0, α', β, a, b, c, d of the other four E_{ref} methods (Table 1) by least squares regression against the simulated daily $E_{\mathrm{ref\_PM}}$, for a specific calibration period. Subsequently, daily values of E_{ref} are calculated for each method during the full period, i.e. 1906-2013, and deviations between the empirical E_{ref} methods and $E_{\mathrm{ref\_PM}}$ are calculated. The sensitivity of E_{ref} to the choice of calibration period is evaluated for each of the methods using $E_{\mathrm{ref\_PM}}$ as a basis.

2.4 Synthetic evaporation series

Synthetic time series of the three evaporation components are derived to systematically unravel the use of empirical crop factors. The synthetic time series are based on the physical model SWAP (Van Dam et al., 2008; Kroes et al., 2009) from which E_{t}, E_{s} and E_{i} can be simulated separately. From these simulations we derive monthly K-values for each E_{ref} method (5x) and vegetation class (5x) (Fig. 2: §2.5), which are subsequently used to derive the corresponding potential evaporation components (5x5x3) using the two-step approach (Fig. 2: §2.6).

Standard values for the vegetation classes and their schematization are taken from the National Hydrologic Instrument (NHI, http://www.nhi.nu/nhi_uk.html; De Lange et al. (2014)) of The Netherlands. The vegetation schematization is constant throughout the period 1906-2013, i.e. dynamic vegetation is not simulated. We consider five natural vegetation classes: grassland (height = 0.5 m and no full soil cover, i.e. not to be confused with the reference grass), heather, deciduous forest, pine forest and spruce forest. Parameters are chosen following NHI (2008) and are provided in the supplementary material. It should be noted that we do not discuss the exact validity of the parameter values used, as we are only concerned with evaporation sensitivity to non-stationary climate within the range of typical vegetation.

SWAP simulates the potential evaporation components of a crop or vegetation class based on the aerodynamic resistance, height, Leaf Area Index (LAI), and albedo. SWAP uses the Penman-Monteith equation, parameterized for each vegetation class to simulate E_t (potential transpiration) and E_s (potential soil evaporation). In case of intercepted precipitation, the values of E_t and E_s are reduced (Van Dam et al., 2008). Interception, which partly evaporates (E_i) and partly drips to the ground, is estimated following Von Hoyningen-Hüne (1983) and Braden (1985) for short vegetation and Gash et al. (1995) for forests. For an extended description of SWAP and the procedures for calculating E_t, E_s and E_i, we refer to Kroes et al. (2009) and Van Dam et al. (2008). Given the international recognition of the SWAP model and successful testing, we assume that the model is able to produce representative synthetic estimates of each evaporation component.

As K_t and K_{ts} are defined for a vegetated surface with a dry canopy (i.e. without interception) and K_{tot} includes interception (see introduction), two different SWAP runs are performed for each vegetation class, without and with interception. Throughout the paper, E_t and E_s are valid for conditions with a dry canopy, whereas E_{tot} includes interception and its limiting effect on transpiration and soil evaporation.

2.5 Derivation of K_t, K_{ts} and K_{tot}

We derive K_t, K_{ts} and K_{tot} for each vegetation class (5x) and E_{ref} method (5x) based on the synthetic E_t, E_s and E_{tot} time series, and the equations given in Table 2. Similar to the calibration of E_{ref} methods, K-values are derived for a specific calibration period, (e.g. 1906-1935). K-values for each vegetation class and E_{ref} method are derived as monthly averages over the calibration period.

2.6 Calculation of potential evaporation components using the two-step approach

Potential evaporation components, E_t, E_t plus E_s (hereafter E_t&E_s) and E_{tot}, for each vegetation class and method are calculated from the daily E_{ref} values by multiplying it with the corresponding K-values, respectively K_t, K_{ts} and K_{tot}, for each vegetation class. Using these three definitions of crop factors separately allows quantifying the error that is made by correcting for each evaporation component.

E_{ref} estimates that are calibrated for a specific period, combined with K-values determined for the same period, are used to calculate daily values of E_t, E_t&E_s and E_{tot} for the full period, i.e. 1906-2013. This procedure corresponds to what is commonly done using the two-step approach, where the empirical parameters of an E_{ref} method are fixed for the region in question, along with the corresponding K-values. Here, we determine the deviation that is potentially introduced when this approach is applied outside its calibration range (period and region/site) in a changing environment, by comparing $E_t(E_{ref},K_t)$, E_t&$E_s(E_{ref},K_{ts})$ and $E_{tot}(E_{ref},K_{tot})$ obtained by the two-step approach with the synthetic 'observed' E_t, E_t&E_s and E_{tot} series.

3 Results

3.1 Calibration period and reference evaporation

Fig. 4 shows the 30-year backwards-looking moving average E_{ref} according to PM, Mak, PT, Har and BC, with the four latter models calibrated to fit the simulated E_{ref_PM} for the first 30-year period, i.e. the calibration period 1906-1935. The minor differences seen between all 30-year mean E_{ref} values during the calibration period (Fig. 4A, year 1935) indicate that each method was calibrated successfully. Using the calibrated equations, E_{ref}'s are calculated for the period 1906-2013, i.e. also outside the calibration period. All empirical models are evaluated with respect to the physically based E_{ref_PM}, which was also used when calibrating the empirical coefficients. The radiation based methods, Mak and PT, deviate only slightly from PM on average with no consistent bias (Fig. 4D), which can be explained by the relatively strong correlations between the trend in 30-year average E_{ref_PM} (Figure 4A) and R_s and R_n (Figure 3F and G); Pearson's $r = 0.85$ and 0.70 for E_{ref_PM} vs. R_s and E_{ref_PM} vs. R_n respectively. ~~, while t~~The temperature based methods, Har and BC, deviate systematically from PM, which can be explained by the relatively weak correlation between the trend in 30 year average $E_{\mathrm{ref\_PM}}$ (Figure 4A) and $\overline{T}$ (Figure 3B); Pearson's $r = 0.17$ for $E_{\mathrm{ref\_PM}}$ vs. $\overline{T}$. This shows that, as the energy used for evaporation mainly comes from direct solar radiation and to a lesser extent from air temperature, temperature based models fail if radiation and temperature trends are weakly correlated (see Figure 3; Pearson's r for 30-year average R_{s} vs. $\overline{T}$ = 0.22). Additionally, Har and BC, ~~each~~ deviate in different directions (Fig. 4B and D): Har consistently underestimates E_{ref}, whereas BC consistently overestimates E_{ref}. This opposite trend is related to the decreasing trend in T_{max}-T_{min} (used in Har), while $\overline{T}$ increases (Figure 3). All four empirical models are unable to reproduce the extreme high evaporation values predicted by PM, especially Har and BC (Fig. 4C). The deviations from $E_{\mathrm{ref\_PM}}$ are considerably larger for individual years (Fig. 4D) than for the 30-year moving average (Fig. 4B).

In practice, 30 year observed time series of evaporation are rarely available for calibration. Therefore, Fig. 5 shows the effect of calibration period length on estimates of E_{ref} for the current climate (1984-2013). This effect is expressed as the maximum absolute deviation of the 30-year average with respect to $E_{\mathrm{ref\_PM}}$. Fig. 5 was compiled by first calibrating the empirical E_{ref} coefficients for all possible calibration periods (in 1906-2013) with a given length (2-30 years) and then simulating E_{ref} for the period 1984-2013 using the calibrated coefficients. The largest deviations occur for shorter calibration periods, as expected. Specific years may cause large deviations when the obtained empirical coefficients are applied outside the calibration period. Deviation decreases notably with increasing calibration periods, suggesting that using more calibration data should result in more stable and accurate E_{ref} estimates. As the calibration period length decreases, deviations in the 30-year average E_{ref} for 1984-2013 increase exponentially.

It should be noted that only deviations in 30-year averages are shown for varying calibration period lengths; deviations in the underlying yearly values are larger, as indicated by Fig. 4D. Additionally, the amplitude of the deviations shown in Fig. 4B and D would increase when calibrated using periods shorter than 30 years (Fig. 5).

3.2 Crop factors and potential evaporation components

Fig. 6 gives monthly average synthetic evaporation components E_t, E_s and E_tot which were used to derive monthly crop factors (three methods: Table 2) for five vegetation classes and five E_ref methods (Table 1), i.e. 3x5x5 crop factors for each calibration period. In contrast to the reference grass surface, the grassland of Fig. 6 does not fully cover the soil, which results in higher E_s and lower E_t. Fig. 7 shows simulated E_t (the two-step approach) for the period 1906-2013, using empirical coefficients for each E_ref method and matching K_t-values, all calibrated on the period 1906-1935. The general patterns in E_t correspond to those of E_ref (Fig. 4B), meaning that the deviations introduced by the two-step approach are mainly determined by the empirical coefficients in the E_ref methods.

The deviation introduced for E_t derived from $E_\mathrm{ref\_PM}$ and $K_\mathrm{t\_PM}$ is relatively minor compared to what is found for the empirical E_ref methods, especially for short vegetation. Apparently, $E_\mathrm{ref\_PM}$ follows the trend in E_t (also obtained using the Penman-Monteith equation, but parameterized for each vegetation class, section 2.4) and the ratio of E_t and $E_\mathrm{ref\_PM}$, used to estimate $K_\mathrm{t\_PM}$, changes little with time for short vegetation. More significant effects of $K_\mathrm{t\_PM}$ are seen for taller vegetation, as climate induced temporal changes in $E_\mathrm{ref\_PM}$ show a height dependent nonlinear relation to changes in E_t (Allen et al., 1998). Therefore, the deviation introduced when using $E_\mathrm{ref\_PM}$ is larger for forests than for the short vegetation classes (Fig. 7). Similar to what is seen in Fig. 4, the deviations for individual years can be considerably larger than the climatic averages.

Fig. 8 shows the sensitivity of crop factors, K, with respect to the calibration period length for heather and spruce forest. The variation in K decreases with increasing calibration length for all methods, but, except for $E_\mathrm{ref\_Mak}$, the variability of K values for the empirical E_ref methods is larger than for $E_\mathrm{ref\_PM}$. These differences are especially notable for forests and illustrate that a poor relationship between the E_ref method and the synthetic potential evaporation component (section 2.4) is compensated by K values that thus show a larger variation over time. Remarkable is the low variability in K_tot values for heather (Fig. 8C), which indicates that the variability seen for K_t (Fig. 8A) is reduced by interception. However, for spruce forest, for which interception is much more dominant, interception increases the variability in K_tot.

From Fig. 8A and D, it can be concluded that the deviations shown in Fig. 7 will increase when shorter calibration periods are used, irrespective of the applied E_ref method. Fig. 9

shows the effect of period (years and length) on the maximum absolute deviation made by the two-step approach for each E_{ref} method and for K_t, K_{ts} and K_{tot}. Fig. 9 confirms that deviations in climatic average evaporation components obtained by applying the two-step approach will generally increase when shorter calibration periods are used. Additionally, Fig. 9 illustrates that deviations are i) larger for tall vegetation than for short vegetation and ii) larger for K_{tot} than for K_t and K_{ts} for vegetation classes with high interception, as is the case for spruce forest. The large deviations for E_{tot} for spruce forest confirm the remark by De Bruin and Lablans (1998), that for wet forest evaporation, the crop factor approach will not be sufficient. Nevertheless, when derived for a sufficiently long time series, the deviations level out and there is no detectable bias.

3.3 Propagation of dimming/brightening periods

In contrast to Fig. 9, which only shows the maximum absolute deviations for the 30-year average potential evaporation components for the years 1984-2013 as a function of the calibration period length, Fig. 10 includes the results of all underlying deviations for heather and spruce using $E_{ref\ PT}$. Fig. 10 demonstrates that climate variability induces systematic overestimation or underestimation of the calculated potential evaporation components, depending on the calibration period used. The sign of the error strongly varies with the calibration period, and the inclusion of a single anomalous year can change the sign of the error.

Fig. 10 further shows that anomalous years or multi-annual climate patterns tend to propagate considerable errors outside the calibration period to the current climate (1984-2013). The patterns of deviations from the synthetic 'observations' show similarities to the global dimming and brightening periods (see Introduction): the first warming period (about 1925-1945) causes a systematic overestimation up to calibration lengths of 30 years, although specific calibration years may result in an underestimation for shorter calibration lengths. The succeeding period of cooling leads to a systematic overestimation, while the second warming period (starting around 1975) results in a more variable pattern. The latter may be linked to the finding of Wang and Dickinson (2013) that, in contrast to the years until the 1970's, there is no significant relationship between variations in temperature and global radiation in following years.

The patterns are comparable for E_t, E_t&E_s and E_{tot}, based on K_t, K_{ts} and K_{tot} respectively, for short vegetation classes. However, for tall vegetation classes with high interception capacity, e.g. spruce (Fig. 10F), using K_{tot} results in a ~~more noisy~~noisier pattern due to specific years of high precipitation. Additionally, including interception may shift the sign of the error.

4 Discussion

4.1 Temporal robustness in hydrological modeling

In this paper we systematically unraveled and quantified how empirical coefficients in the two-step approach affect estimates of potential evaporation. We used the past century's time series of observed climate containing non-stationary signals of multi-decadal atmospheric oscillations, global warming, and global dimming/brightening (Suo et al., 2013; Stanhill, 2007; Wild, 2009; Wild et al., 2005) to evaluate the sensitivity of the two-step approach to both the length of the reference calibration period and the reference years. To this end we calibrated the empirical coefficients of the two-step approach based on different periods and ~~then~~ showed that using the thus obtained empirical coefficients outside their calibration range may lead to systematic differences between E_{ref}-methods, and to systematic errors in estimated potential E components. The signs of the errors for calculated climatic average evaporation components differ, depending on the E_{ref} method used, and on the specific period (length and years) of calibration. Hooghart and Lablans (1988) stated that, for the two-step approach, the correctness of empirical coefficients for the estimation of E_{ref} are of minor importance, as these are compensated by K. However, here we have shown that while this may be true within the calibration period, this statement does not hold when extrapolating. As potential evaporation is a key input in hydrological models, input errors will propagate to estimates of related processes, such as the soil moisture budget, droughts, recharge and groundwater processes.

These results are important because ~~Although this result may seem trivial,~~ the two-step approach, including extrapolating empirical coefficients, is ~~common practice~~frequently applied in hydrological modeling studies, as mentioned in the introduction. Ehret et al. (2014) state that "in hydrological modeling, it is often conveniently assumed that the variables presenting climate vary in time while the general model structure and model parameters representing catchment characteristics remain time-invariant". There is a clear parallel of this statement with the approach presented herein where meteorological conditions vary in time, while climate-dependent (empirical) parameters are often fixed values.

In practice, long time series of observed evaporation are rare and not evenly distributed spatially. As such, for many applications, hydrologists must rely on incomplete calibration data, use analogous stations with similar characteristics, or simply default to published values for crop factors and E_{ref} model parameters. Such published values for empirical factors of the different E_{ref} methods (Table 1) are C_1=0.65, C_0=0 (De Bruin, 1987), α'=1.3, β=0 (De Bruin and Lablans, 1998), a=0.0023, b=17.8 (Droogers and Allen, 2002), c=0, d=1 (Sperna Weiland et al., 2012). Besides absolute values of the calibrated empirical factors, our analysis provides insight into the sensitivity of the results, i.e. the parameter values, to the calibration period. Calibrated model parameters for 30 year calibration periods are (standard deviations between brackets): C_1=0.64(0.01), C_0=0.37(0.03), α'=1.06(0.01), β=0.57(0.02), a=0.0022(0.002), b=20.7(3.0), and c=-2.31(0.06), d=1.73(0.03). For a more realistic calibration period of e.g. three years, the standard deviations increase by a factor 2-10, depending on the E_{ref} method used.

For the Netherlands, published crop factors (K_{tot} for $E_{\mathrm{ref\_Mak}}$ with coefficients C_1 = 0.65 and C_0 = 0) are 1.0, 0.8, 1.1, 1.2 and 1.3 for grass, heather, deciduous forest, pine forest and spruce forest, respectively (Spieksma et al., 1996). Values from our study for e.g. the mean 30 year K_{tot} values, were 0.8, 1.03, 1.02, 1.07 and 1.25, respectively. Problems arise on the applicability of the published and frequently re-used crop factors, as the climatic conditions used for fitting are rarely documented. The analysis herein provides insight in the uncertainty ranges that could be expected using published empirical coefficients (Figure 8) and their potential impact on simulated potential evaporation components.

This study has shown that potential evaporation estimates are most accurate and stable with a long calibration period. However, even when using a long observed record, estimates may include errors due to the assumption of constant empirical coefficients in a non-stationary climate, i.e. the calibration period not being representative of current conditions. Evaporation estimates outside the calibration period are even more susceptible to non-stationarity when the calibration period is relatively short, as with areas where observed evaporation data are sparse. Finally, estimating evaporation based on published typical values without calibration is most susceptible to errors, as these parameters are typically global averages but also contain the non-stationary reference period issues identified in this paper. To remove bias by systematic input errors, as in e.g. evaporation, it is common practice to tune models by calibration (Ehret et al. (2014) and references therein). Although model calibration may compensate for biased input data, resulting in more accurate results and comparable model efficiencies, such calibration limits the general applicability of models when the bias is not constant over time (Andréassian et al., 2004). Figs. 4 and 7 show that such non-constant bias occurs for both E_{ref} and potential E estimates, thus limits their application outside the calibration range.

Although extrapolations to future periods will always include uncertainty, it is important to quantify and limit this uncertainty. This analysis provides such a quantification, identifying the sensitivity of evaporation estimates to extrapolation and representing information on ways to reduce potential errors; e.g. Figure 7 quantifies the error that is made in extrapolations. A similar modeling approach as presented here could be used to identify climate induced changes in potential evaporation components. Moreover, such information can be used to reduce potential errors, if the errors can be explained from e.g. differences in climatic conditions between the periods of calibration and application. We did so for the errors identified for potential evaporation in Figure 7, and found that for all E_{ref} methods except for Har, and for all vegetation classes, the error correlates well with differences in relative humidity ($R^2 > 0.78$) (Figure 3). This makes the errors predictable, and provides opportunities to correct for them.

Although we advocate using process-based evaporation simulations where possible, it should be emphasized that the two-step approach still can be a valuable concept, especially in regions with limited data availability. However, some considerations may strengthen the robustness of the two-step approach. First, our results show that applying radiation based methods are preferred over temperature based methods. Second, ideally, independent of the type of empirical method used, the coefficients should be recalibrated against measurements. Third, as such recalibration will practically often not be feasible, we advocate to identify changes in climatic conditions for the period of application and the calibration period, and to quantify using a sensitivity analysis, how they may impact potential evaporation estimates. This provides uncertainty ranges that advance the interpretation of modeling exercises.

4.2 ~~Propagation of dimming/brightening periods~~

~~In contrast to Fig. 9, which only shows the maximum absolute deviations for the 30-year average potential evaporation components for the years 1984-2013 as a function of the calibration period length, Fig. 10 includes the results of all underlying deviations for heather and spruce using E_{ref_PT}. Fig. 10 demonstrates that climate variability induces systematic overestimation or underestimation of the calculated potential evaporation components, depending on the calibration period used. The sign of the error strongly varies with the calibration period, and the inclusion of a single anomalous year can change the sign of the error.~~

~~Fig. 10 further shows that anomalous years or multi-annual climate patterns tend to propagate considerable errors outside the calibration period to the current climate (1984-2013). The patterns of deviations from the synthetic 'observations' show similarities to the global dimming and brightening periods (see Introduction): the first warming period (about 1925-1945) causes a systematic overestimation up to calibration lengths of 30 years, although specific calibration years may result in an underestimation for shorter calibration lengths. The succeeding period of cooling leads to a systematic overestimation, while the second warming period (starting around 1975) results in a more variable pattern. The latter may be linked to the finding of Wang and Dickinson (2013) that, in contrast to the years until the 1970's, there is no significant relationship between variations in temperature and global radiation in following years.~~

~~The patterns are comparable for E_t, E_t&E_s and E_{tot}, based on K_t, K_{ts} and K_{tot} respectively, for short vegetation classes. However, for tall vegetation classes with high interception capacity, e.g. spruce (Fig. 10F), using K_{tot} results in a more noisy pattern due to specific years of high precipitation. Additionally, including interception may shift the sign of the error.~~

4.3 Implications for climate change impact studies

Poor transferability of parameter estimates made during calibration can have potentially large impacts for studies in non-stationary conditions (Coron et al., 2014), e.g. for climate change impact studies (Bormann, 2011; Karlsson et al., 2014). To improve the temporal robustness of hydrological modeling, Coron et al. (2014) propose, while ~~putting~~ adding it to~~in~~ the framework of the new IAHS Scientific Decade "Panta Rhei" (Montanari et al., 2013), to

~~particularly~~ especially advance ~~in the~~our abilit~~y~~ies to estimate temporal variations in evaporation fluxes. This study contributes to this larger objective.

For climate change impact studies, applications of empirical models are particularly problematic, as empirical methods closely approximate observations of natural processes, but do not capture the underlying physics. When extrapolating to new climate regimes, these assumptions are not guaranteed to remain valid (Kay and Davies, 2008; Bormann, 2011; Arnell, 1999). Similar to our findings, simulating historic non-stationary climatic conditions, Kay and Davies (2008) demonstrate that $E_{\mathrm{ref\_PM}}$ and temperature based E_{ref} methods give different projected evaporation estimates when applied to future climate model data. Additionally, Haddeland et al. (2011) show, using the WATCH climate forcing data (Weedon et al., 2011), that global hydrological models that differ in their choice of evaporation schemes, show significantly different evaporation estimates. These large discrepancies in an important part of the water cycle may have a large effect on the modeled hydrological impacts of climate change and increases the uncertainty of impact estimates (Bormann, 2011; Kay and Davies, 2008; Haddeland et al., 2011).

To show the implications of using different empirical E_{ref} methods in hydrological applications under recent climate change, without the need for numerous extensive model runs, we calculated the Standardized Precipitation and Evaporation Index (SPEI) for the period 1906-2013, with the empirical coefficients calibrated for the 30-year period 1906-1935.

The SPEI (Beguería et al., 2013; Vicente-Serrano et al., 2010) is a commonly used meteorological drought index, which is a variant of the WMO-recommended Standardized Precipitation Index SPI (Guttman, 1999; Hayes et al., 2011; McKee et al., 1993). Unlike the SPI, which calculates precipitation accumulated over a period and then normalizes the accumulated value based on typical seasonal conditions, the SPEI instead normalizes the accumulated difference of the climatic water balance, defined as the difference between precipitation and E_{ref}. This produces a time series of normalized values, such that an SPEI of 0 refers to typical conditions, an SPEI of negative one refers to a condition where $\Sigma(P - E_{\mathrm{ref}})$ is one standard deviation drier than typical, and vice versa for positive one. For this example, the SPEI6 was calculated, normalizing the climatic water balance summed over the preceding six months, following the fitting procedures outlined in Stagge et al. (2014a) and Gudmundsson and Stagge (2014).

Fig. 11~~11~~ shows the results of this analysis, with the assumed accurate SPEI6, based on E_{ref_PM}, shown at the top and the difference between this and SPEI6 for all other empirical reference evaporation models shown below. As with the results of E_{ref} simulations, the Mak and PT models are closest to the observed signal (differences in the range of -0.2 to 0.2), while the Har and BC models produce greater variability (ΔSPEI6 = -0.5 to 0.5). Differences of this magnitude can make a large difference when interpreting drought risk. For example, the year 1947 produced a severe drought at the De Bilt site (SPEI6 = -2.2); however all other methods underestimate E_{ref}, producing SPEI6 values between -1.5 and -1.9. This in turn, changes the interpretation of this drought from an event expected to occur once every 72 years to an event expected to occur once every 15-35 years. This is a significant difference in risk level which can be attributed to differences among the evaporation methods and a potentially non-representative calibration period. SPEI sensitivity to E_{ref} method is analyzed in greater detail in Stagge et al. (2014b).

5 Conclusion

In this study we thoroughly analyzed the robustness of the two-step approach to simulate potential evaporation. We ~~show that~~quantified the magnitude of the systematic errors that may be introduced when empirical coefficients are applied outside their calibration period, ~~and that~~depending on ~~the magnitude of these errors depends on~~ differences in climate, the period, and the length of the calibration period. Our hydrological models are to varying extent regression models, which limits their general applicability, and the estimation of potential evaporation is closely linked to climate variability. With our analysis, we want to raise awareness and to provide a quantification of possible systematic errors that may be introduced in estimates of potential evaporation and in hydrological modeling studies due to straightforward application of i) the common two-step approach for potential evaporation specifically, and ii) fixed instead of time-variant model parameters in general.

6 Acknowledgements

This study was carried out under the auspices of the joint research program of the Dutch Water Utility sector. We would like to thank the three anonymous reviewers for their valuable suggestions to improve the manuscript.

References

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E., and Rasmussen, J.: An introduction to the European Hydrological System -- Systeme Hydrologique Europeen, "SHE", 2: Structure of a physically-based, distributed modelling system, J Hydrol, 87, 61-77, 1986.

Allen, R. G., and Pruitt, W. O.: Rational use of the FAO Blaney-Criddle formula, Journal of Irrigation and Drainage Engineering, 112, 139-155, 1986.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration - Guidelines for computing crop water requirements FAO Irrigation and drainage paper, FAO - Food and Agriculture Organization of the United Nations, Rome, 1998.

Allen, R. G.: Using the FAO-56 dual crop coefficient method over an irrigated region as part of an evapotranspiration intercomparison study, J Hydrol, 229, 27-41, 2000.

Allen, R. G., Pereira, L. S., Smith, M., Raes, D., and Wright, J. L.: FAO-56 Dual crop coefficient method for estimating evaporation from soil and application extensions, Journal of Irrigation and Drainage Engineering, 1, 2-13, 2005.

Allen, R. G., Pereira, L. S., Howell, T. A., and Jensen, M. E.: Evapotranspiration information reporting: I. Factors governing measurement accuracy, Agric Water Manag, 98, 899-920, http://dx.doi.org/10.1016/j.agwat.2010.12.015, 2011.

Andréassian, V., Perrin, C., and Michel, C.: Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models, J Hydrol, 286, 19-35, 2004.

Arnell, N. W.: The effect of climate change on hydrological regimes in Europe: a continental perspective, Global Environmental Change, 9, 5-23, http://dx.doi.org/10.1016/S0959-3780(98)00015-6, 1999.

Beguería, S., Vicente-Serrano, S. M., Reig, F., and Latorre, B.: Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring, International Journal of Climatology, doi: 10.1002/joc.3887, 2013.

Bergström, S., and Forsman, A.: Development of a conceptual deterministic rainfall-runoff model, Nord Hydrol, 4, 147-170, 1973.

Blaney, H. F., and Criddle, W. P.: Determining water requirements in irrigated areas from climatological and irrigation data, USDA (SCS), 48, 1950.

Bormann, H.: Sensitivity analysis of 18 different potential evapotranspiration models to observed climatic change at German climate stations, Clim Change, 104, 729-753, 10.1007/s10584-010-9869-7, 2011.

Braden, H.: Ein Energiehaushalts- und Verdunstungsmodell for Wasser und Stoffhaushaltsuntersuchungen landwirtschaftlich genutzer Einzugsgebiete, Mittelungen Deutsche Bodenkundliche Geselschaft, 42, 294-299, 1985.

Bradford, R. B., Ragab, R., Crooks, S. M., Bouraoui, F., and Peters, E.: Simplicity versus complexity in modelling groundwater recharge in Chalk catchments, Hydrology and Earth System Sciences, 6, 927-937, 10.5194/hess-6-927-2002, 1999.

Calder, I. R.: Assessing the water use of short vegetation and forests: Development of the Hydrological Land Use Change (HYLUC) model, Wat Resour Res, 39, 1318, 10.1029/2003wr002040, 2003.

Coron, L., Andréassian, V., Perrin, C., Bourqui, M., and Hendrickx, F.: On the lack of robustness of hydrologic models regarding water balance simulation: a diagnostic approach applied to three models of increasing complexity on 20 mountainous catchments, Hydrology and Earth System Sciences, 18, 727-746, 10.5194/hess-18-727-2014, 2014.

De Bruin, H. A. R., and Holtslag, A. A. M.: A simple parameterization of the surface fluxes of sensible and latent heat during daytime compared with the Penman-Monteith concept, Journal of Applied Meteorology, 21, 1610-1621, 10.1175/1520-0450(1982)021<1610:aspots>2.0.co;2, 1982.

De Bruin, H. A. R.: From Penman to Makkink, in: Evaporation and weather, edited by: Hooghart, C., Proceedings and Information. Comm. Hydrological Research TNO, The Hague, 5-32, 1987.

De Bruin, H. A. R., and Lablans, W. N.: Reference crop evapotranspiration determined with a modified Makkink equation, Hydrological Processes, 12, 1053-1062, 1998.

De Lange, W. J., Prinsen, G. F., Hoogewoud, J. C., Veldhuizen, A. A., Verkaik, J., Oude Essink, G. H. P., van Walsum, P. E. V., Delsman, J. R., Hunink, J. C., Massop, H. T. L., and Kroon, T.: An operational, multi-scale, multi-model system for consensus-based, integrated water management and policy analysis: The Netherlands Hydrological Instrument, Environmental Modelling & Software, 59, 98-108, http://dx.doi.org/10.1016/j.envsoft.2014.05.009, 2014.

Doorenbos, J., and Pruitt, W. O.: Crop water requirements; Irrigation and drainage paper No. 24, FAO, Rome, 1977.

Driessen, T. L. A., Hurkmans, R. T. W. L., Terink, W., Hazenberg, P., Torfs, P. J. J. F., and Uijlenhoet, R.: The hydrological response of the Ourthe catchment to climate change as modelled by the HBV model, Hydrology and Earth System Sciences, 14, 651-665, 10.5194/hess-14-651-2010, 2010.

Droogers, P., and Allen, R. G.: Estimating reference evapotranspiration under inaccurate data conditions, Irrigation and Drainage Systems, 16, 33-45, 10.1023/a:1015508322413, 2002.

Ehret, U., Gupta, H. V., Sivapalan, M., Weijs, S. V., Schymanski, S. J., Blöschl, G., Gelfan, A. N., Harman, C., Kleidon, A., Bogaard, T. A., Wang, D., Wagener, T., Scherer, U., Zehe, E., Bierkens, M. F. P., Di Baldassarre, G., Parajka, J., van Beek, L. P. H., van Griensven, A., Westhoff, M. C., and Winsemius, H. C.: Advancing catchment hydrology to deal with predictions under change, Hydrology and Earth System Sciences, 18, 649-671, 10.5194/hess-18-649-2014, 2014.

Farmer, W., Strzepek, K., Schlosser, C. A., Droogers, P., and Gao, X.: A method for calculating reference evapotranspiration on daily time scales, MIT Joint Program on the Science and Policy of Global Change, 2011.

Feddes, R. A.: Crop factors in relation to Makkink reference-crop evapotranspiration, in: Evaporation and weather, edited by: Hooghart, C., Proceedings and Information. Comm. Hydrological Research TNO, The Hague, 33-47, 1987.

Feddes, R. A., and Lenselink, K. J.: Evapotranspiration, in: Drainage principles and applications, edited by: Ritzema, H. P., International Institute for Land Reclamation and Improvement ILR, Wageningen, 1994.

Federer, C. A., Vörösmarty, C., and Fekete, B.: Intercomparison of methods for calculating potential evaporation in regional and global water balance models, Water Resour. Res., 32, 2315-2321, 10.1029/96wr00801, 1996.

Gash, J. H. C., Lloyd, C. R., and Lachaud, G.: Estimating sparse forest rainfall interception with an analytical model, J Hydrol, 170, 79-86, 1995.

Gudmundsson, L., and Stagge, J. H.: SCI: Standardized Climate Indices such as SPI, SRI or SPEI. R package version 1.0-1, 2014.

Guerschman, J. P., Van Dijk, A. I. J. M., Mattersdorf, G., Beringer, J., Hutley, L. B., Leuning, R., Pipunic, R. C., and Sherman, B. S.: Scaling of potential evapotranspiration with MODIS data reproduces flux observations and catchment water balance observations across Australia, J Hydrol, 369, 107-119, http://dx.doi.org/10.1016/j.jhydrol.2009.02.013, 2009.

Guttman, N. B.: Accepting the Standardized Precipitation Index: A calculation algorithm, in, Wiley Online Library, 1999.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel estimate of the global terrestrial water balance: Setup and first results, Journal of Hydrometeorology, 12, 869-884, 10.1175/2011jhm1324.1, 2011.

Hayes, M., Svoboda, M., Wall, N., and Widhalm, M.: The Lincoln declaration on drought indices: universal meteorological drought index recommended, Bull Am met Soc, 92, 485-488, doi: 10.1175/2010BAMS3103.1, 2011.

Hooghart, J. C., and Lablans, W. N.: van Penman naar Makkink: een nieuwe berekeningswijze voor de klimatologische verdampingsgetallen, KNMI, De Bilt, 1988.

Hupet, F., and Vanclooster, M.: Effect of the sampling frequency of meteorological variables on the estimation of the reference evapotranspiration, J Hydrol, 243, 192-204, http://dx.doi.org/10.1016/S0022-1694(00)00413-3, 2001.

Hurkmans, R. T. W. L., Terink, W., Uijlenhoet, R., Moors, E. J., Troch, P. A., and Verburg, P. H.: Effects of land use changes on streamflow generation in the Rhine basin, Wat Resour Res, 45, W06405, 10.1029/2008wr007574, 2009.

Jones, P. D., and Moberg, A.: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001, Journal of Climate, 16, 2003.

Karlsson, I. B., Sonnenborg, T. O., Jensen, K. H., and Refsgaard, J. C.: Historical trends in precipitation and stream discharge at the Skjern River catchment, Denmark, Hydrology and Earth System Sciences, 18, 595-610, 10.5194/hess-18-595-2014, 2014.

Kay, A. L., and Davies, H. N.: Calculating potential evaporation from climate model data: A source of uncertainty for hydrological climate change impacts, J Hydrol, 358, 221-239, http://dx.doi.org/10.1016/j.jhydrol.2008.06.005, 2008.

Kay, A. L., Bell, V. A., Blyth, E. M., Crooks, S. M., Davies, H. N., and Reynard, N. S.: A hydrological perspective on evaporation: historical trends and future projections in Britain, Journal of Water and Climate Change, 4, 193-208, doi:10.2166/wcc.2013.014, 2013.

Klein Tank, A. M. G., and Lenderink, G.: Klimaatverandering in Nederland; Aanvullingen op de KNMI'06 scenario's, KNMI, De Bilt, 2009.

Kroes, J. G., Van Dam, J. C., Groenendijk, P., Hendriks, R. F. A., and Jacobs, C. M. J.: SWAP version 3.2, Theory description and user manual, Wageningen University and Research Centre, WageningenAlterra report 1649 (update 02), 2009.

Lenderink, G., Buishand, A., and van Deursen, W.: Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach, Hydrology and Earth System Sciences, 11, 1145-1159, 10.5194/hess-11-1145-2007, 2007.

Makkink, G. F.: Testing the Penman formula by means of lysimeters, Journal of the Institution of Water Engineers, 11, 277-288, 1957.

McKee, T. B., Doesken, N. J., and Kleist, J.: The relationship of drought frequency and duration to time scales, Proceedings of the 8th Conference on Applied Climatology, 1993, 179-183,

Montanari, A., Young, G., Savenije, H., Hughes, D., Wagener, T., Ren, L., Koutsoyiannis, D., Cudennec, C., Toth, E., and Grimaldi, S.: "Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022, Hydrological Sciences Journal, 58, 1256-1275, 2013.

Monteith, J.: Evaporation and environment, Symp. Soc. Exp. Biol, 1965, 4,

NHI, N. H. I.: Modelrapportage; Deelrapport Gewaskenmerken. Versie: NHI\FASE_1+\2008\DR12\v2 2008.

Penman, H. L.: Natural evaporation from open water, bare soil and grass, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 193, 120-145, 10.1098/rspa.1948.0037, 1948.

Penman, H. L.: Evaporation: An introductory survey, Neth J Agric Sci, 4, 9-29, 1956.

Sanchez-Lorenzo, A., Calbó, J., and Martin-Vide, J.: Spatial and temporal trends in sunshine duration over Western Europe (1938–2004), Journal of Climate, 21, 6089-6098, 10.1175/2008jcli2442.1, 2008.

Savenije, H. H. G.: The importance of interception and why we should delete the term evapotranspiration from our vocabulary, Hydrological Processes, 18, 1507-1511, 10.1002/hyp.5563, 2004.

Seibert, J.: Reliability of model predictions outside calibration conditions, Nord Hydrol, 34, 477-492, 10.2166/nh.2003.028, 2003.

Shabalova, M. V., Deursen, W. P. A. v., and Buishand, T. A.: Assessing future discharge of the river Rhine using regional climate model integrations and a hydrological model, Clim Res, 23, 233-246, 10.3354/cr023233, 2003.

Shaw, S. B., and Riha, S. J.: Assessing temperature-based PET equations under a changing climate in temperate, deciduous forests, Hydrological Processes, 25, 1466-1478, 10.1002/hyp.7913, 2011.

Shuttleworth, W. J.: Putting the 'vap' into evaporation, Hydrology and Earth System Sciences, 11, 210-244, 2007.

Solomon, S., Qin, D., Manning, M., Alley, R. B., Berntsen, T., Bindoff, N. L., Chen, Z., Chidthaisong, A., Gregory, J. M., Hegerl, G. C., Heimann, M., Hewitson, B., Hoskins, B. J., Joos, F., Jouzel, J., Kattsov, V., Lohmann, U., Matsuno, T., Molina, M., Nicholls, N., Overpeck, J., Raga, G., Ramaswamy, V., Ren, J., Rusticucci, M., Somerville, R., Stocker, T. F., Whetton, P., Wood, R. A., and Wratt, D.: Technical Summary, in: Climate Change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Sperna Weiland, F. C., Tisseuil, C., Dürr, H. H., Vrac, M., and van Beek, L. P. H.: Selecting the optimal method to calculate daily global reference potential evaporation from CFSR reanalysis data for application in a hydrological model study, Hydrology and Earth System Sciences, 16, 983-1000, 10.5194/hess-16-983-2012, 2012.

Spieksma, J. F. M., Dolman, A. J., and Schouwenaars, J. M.: De parameterisatie van de verdamping van natuurterreinen in hydrologische modellen, 1996.

Stagge, J. H., Tallaksen, L. M., L., G., Van Loon, A. F., and Stahl, K.: Pan-European comparison of candidate distributions for climatological drought indices (SPI and SPEI), International Journal of Climatology, In Review, 2014a.

Stagge, J. H., Tallaksen, L. M., Xu, C.-Y., and Van Lanen, H. A. J.: Standardized precipitation-evapotranspiration index (SPEI): Sensitivity to potential evapotranspiration model and parameters. Proceedings of FRIEND-Water 2014 (IAHS Publ. 363, 2014). Montpellier, France, October 2014., 2014b,

Stanhill, G.: A perspective on global warming, dimming, and brightening, Eos, Transactions American Geophysical Union, 88, 58-58, 10.1029/2007eo050007, 2007.

Suo, L., Otterå, O. H., Bentsen, M., Gao, Y., and Johannessen, O. M.: External forcing of the early 20th century Arctic warming, Tellus A, 65, 2013.

Tiktak, A., and Bouten, W.: Soil water dynamics and long-term water balances of a Douglas fir stand in the Netherlands, J Hydrol, 156, 265-283, http://dx.doi.org/10.1016/0022-1694(94)90081-7, 1994.

Trambauer, P., Dutra, E., Maskey, S., Werner, M., Pappenberger, F., van Beek, L. P. H., and Uhlenbrook, S.: Comparison of different evaporation estimates over the African continent, Hydrology and Earth System Sciences, 18, 193-212, 10.5194/hess-18-193-2014, 2014.

Van Dam, J. C., Groenendijk, P., Hendriks, R. F. A., and Kroes, J. G.: Advances of modeling water flow in variably saturated soils with SWAP, Vadose Zone Journal, 7, 640-653, DOI:10.2136/vzj2007.0060, 2008.

Van den Hurk, B., Klein Tankink, A., Lenderink, G., Van Ulden, A., Van Oldenborgh, G. J., Katsman, C., Van den Brink, H., Keller, F., Bessembinder, J., Burgers, G., Komen, G., Hazeleger, W., and Drijfhout, S.: KNMI Climate change scenarios 2006 for the Netherlands, KNMI, De Bilt, 82 pp., 2006.

Van Roosmalen, L., Sonnenborg, T. O., and Jensen, K. H.: Impact of climate and land use change on the hydrology of a large-scale agricultural catchment, Water Resour. Res., 45, doi:10.1029/2007WR006760, 2009.

Van Walsum, P. E. V., and Supit, I.: Influence of ecohydrologic feedbacks from simulated crop growth on integrated regional hydrologic simulations under climate scenarios, Hydrology and Earth System Sciences, 16, 1577-1593, 10.5194/hess-16-1577-2012, 2012.

Vázquez, R. F., and Feyen, J.: Effect of potential evapotranspiration estimates on effective parameters and performance of the MIKE SHE-code applied to a medium-size catchment, J Hydrol, 270, 309-327, 2003.

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index, Journal of Climate, 23, 1696-1718, 2010.

Von Hoyningen-Hüne, J.: Die Interception des Niederschlags in landwirtschaftlichen Beständen, 1983.

Wallace, J. S.: Calculating evaporation: resistance to factors, Agric For Meteorol, 73, 353-366, DOI:10.1016/0168-1923(94)05084-J, 1995.

Wang, K., and Dickinson, R. E.: Contribution of solar radiation to decadal temperature variability over land, Proceedings of the National Academy of Sciences, 110, 14877-14882, 10.1073/pnas.1311433110, 2013.

Ward, P. J., Renssen, H., Aerts, J. C. J. H., van Balen, R. T., and Vandenberghe, J.: Strong increases in flood frequency and discharge of the River Meuse over the late Holocene: impacts of long-term anthropogenic land use change and climate variability, Hydrology and Earth System Sciences, 12, 159-175, 10.5194/hess-12-159-2008, 2008.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the Twentieth Century, Journal of Hydrometeorology, 12, 823-848, 10.1175/2011jhm1369.1, 2011.

Wild, M., Gilgen, H., Roesch, A., Ohmura, A., Long, C. N., Dutton, E. G., Forgan, B., Kallis, A., Russak, V., and Tsvetkov, A.: From dimming to brightening: Decadal changes in solar radiation at earth's surface, Science, 308, 847-850, 10.1126/science.1103215, 2005.

Wild, M.: Global dimming and brightening: A review, Journal of Geophysical Research: Atmospheres, 114, D00D16, 10.1029/2008jd011470, 2009.

Witte, J. P. M., Runhaar, J., van Ek, R., van der Hoek, D. C. J., Bartholomeus, R. P., Batelaan, O., van Bodegom, P. M., Wassen, M. J., and van der Zee, S. E. A. T. M.: An ecohydrological sketch of climate change impacts on water and natural ecosystems for the Netherlands: bridging the gap between science and society, Hydrology and Earth System Sciences, 16, 3945-3957, 10.5194/hess-16-3945-2012, 2012.

Yamanouchi, T.: Early 20th century warming in the Arctic: A review, Polar Science, 5, 53-71, http://dx.doi.org/10.1016/j.polar.2010.10.002, 2011.

Zhou, M. C., Ishidaira, H., Hapuarachchi, H. P., Magome, J., Kiem, A. S., and Takeuchi, K.: Estimating potential evapotranspiration using Shuttleworth-Wallace model and NOAA-AVHRR NDVI data to feed a distributed hydrological model over the Mekong River basin, J Hydrol, 327, 151-173, 2006.

Tables

Table 1: Equations used to calculate daily values of reference evaporation E_{ref} [mm/d] for the period 1906-2013 at De Bilt meteorological station.

Abbreviation	Method	Equation
$E_{\mathrm{ref\_PM}}$	Penman-Monteith (Monteith, 1965)	$E=\frac{1}{\lambda}\left(\frac{\Delta\left(R_n-G\right)+\rho_a c_p\frac{\left(e_s-e_a\right)}{r_a}}{\Delta+\gamma\left(1+\frac{r_s}{r_a}\right)}\right)$
$E_{\mathrm{ref\_Mak}}$	Makkink (Makkink, 1957)	$E=\frac{1}{\lambda}\left(C_1\frac{\Delta}{\Delta+\gamma}R_s+C_0\right)$
$E_{\mathrm{ref\_PT}}$	Modified Priestley-Taylor (De Bruin and Holtslag, 1982)	$E=\frac{1}{\lambda}\left(\alpha'\frac{\Delta\left(R_n-G\right)}{\Delta+\gamma}+\beta\right)$
$E_{\mathrm{ref\_Har}}$	Hargreaves (Droogers and Allen, 2002; Farmer et al., 2011)	$E=\frac{1}{\lambda}\left(aR_a\left(\bar{T}+b\right)\left(T_{\max}-T_{\min}\right)^{0.5}\right)$
$E_{\mathrm{ref\_BC}}$	Blaney and Criddle (1950) as in Allen and Pruitt (1986)	$E=\frac{1}{\lambda}\left(c+d\left(p\left(0.46\bar{T}+8.13\right)\right)\right)$

Where E = evaporation [kg d^{-1} m^{-2}], λ = latent heat of vaporization [MJ kg^{-1}], Δ = slope of the vapor pressure curve [kPa °C^{-1}], R_a = extraterrestrial radiation [MJ m^{-2} day^{-1}], R_s = solar radiation [MJ m^{-2} day^{-1}], R_n = net radiation [MJ m^{-2} day^{-1}], G = soil heat flux [MJ m^{-2} day^{-1}], ρ_a = mean air density [kg m^{-3}], c_p = specific heat of the air [MJ kg^{-1} °C^{-1}], γ = psychrometric constant [kPa °C^{-1}], r_s = surface resistance [s m^{-1}], r_a = aerodynamic resistance [s m^{-1}], $(e_s - e_a)$ = saturation vapour pressure deficit [kPa], $\bar{T}$, $T_{\max}$ and $T_{\min}$ = mean, maximum and minimum temperature [°C], p = mean daily percentage of annual daytime hours [%]. $\boldsymbol{C_1}$, $\boldsymbol{C_0}$, $\boldsymbol{\alpha'}$, $\boldsymbol{\beta}$, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{d}$ are the coefficients adjusted in the calibration.

Table 2: Equations used to calculate monthly average crop factors for each vegetation class and E_{ref} method.

Crop factor	Description	Equation
$K_{\mathrm{t}}(E_{\mathrm{ref}})$	crop factor for potential transpiration	$K_{\mathrm{t}} = E_{\mathrm{t}} / E_{\mathrm{ref}}$
$K_{\mathrm{ts}}(E_{\mathrm{ref}})$	crop factor for potential transpiration + soil evaporation	$K_{\mathrm{ts}} = \left(E_{\mathrm{t}} + E_{\mathrm{s}}\right) / E_{\mathrm{ref}}$
$K_{\mathrm{tot}}(E_{\mathrm{ref}})$	crop factor for total evaporation	$K_{\mathrm{tot}} = E_{\mathrm{tot}} / E_{\mathrm{ref}}$

Figures

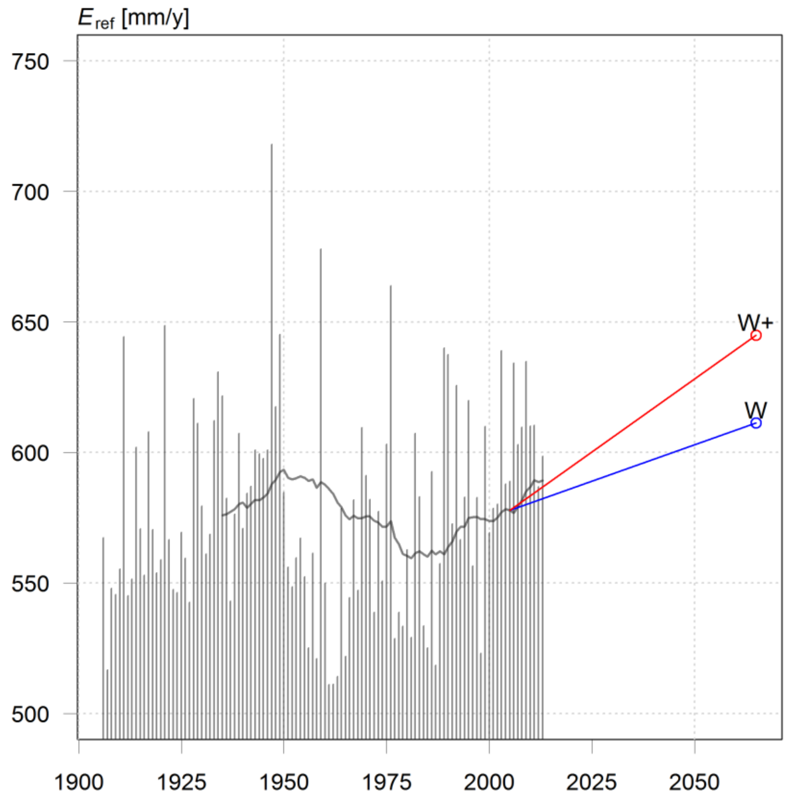

Figure 1: Yearly and 30-year moving average E_{ref} according to Penman-Monteith for De Bilt, the Netherlands and projected E_{ref} values for the period 2036-2065. Projections are based on national climate scenarios (Van den Hurk et al., 2006) developed by the Royal Netherlands Meteorological Institute (KNMI). Two of the scenarios have been found to be most likely (Klein Tank and Lenderink, 2009) and are presented here: scenario W (blue line) and W+ (red line). Both comprise a +2K global temperature increase, but with respectively unchanged and changed (+) air circulation patterns in summer and winter. The scenarios were used to transfer the climatic conditions of 1976-2005 to the period 2036-2065 (Van den Hurk et al., 2006).

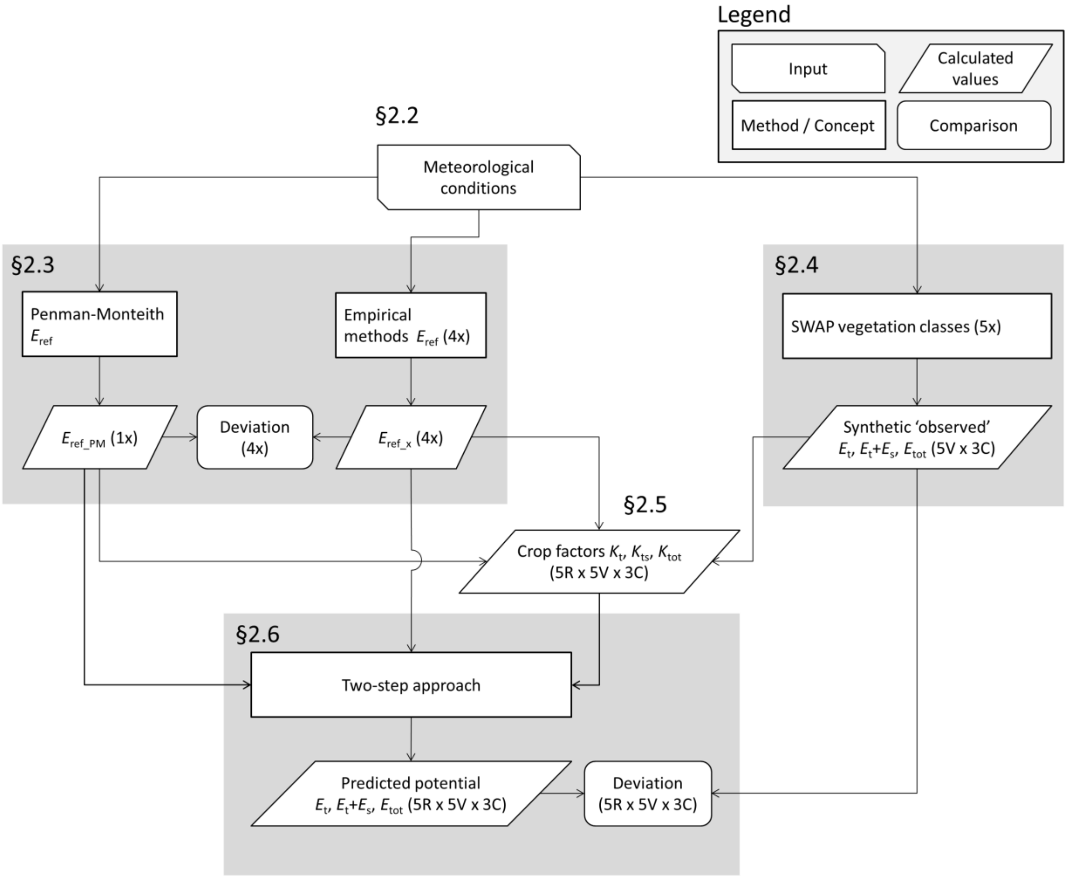

Figure 2: Flow chart of the methodology followed. E_{ref_x} = E_{ref} of the empirical methods Mak, Har, PT and BC; R = number of reference evaporation methods, V = number of vegetation classes, C = number of evaporation components. For the explanation of the other abbreviations we refer to the introduction.

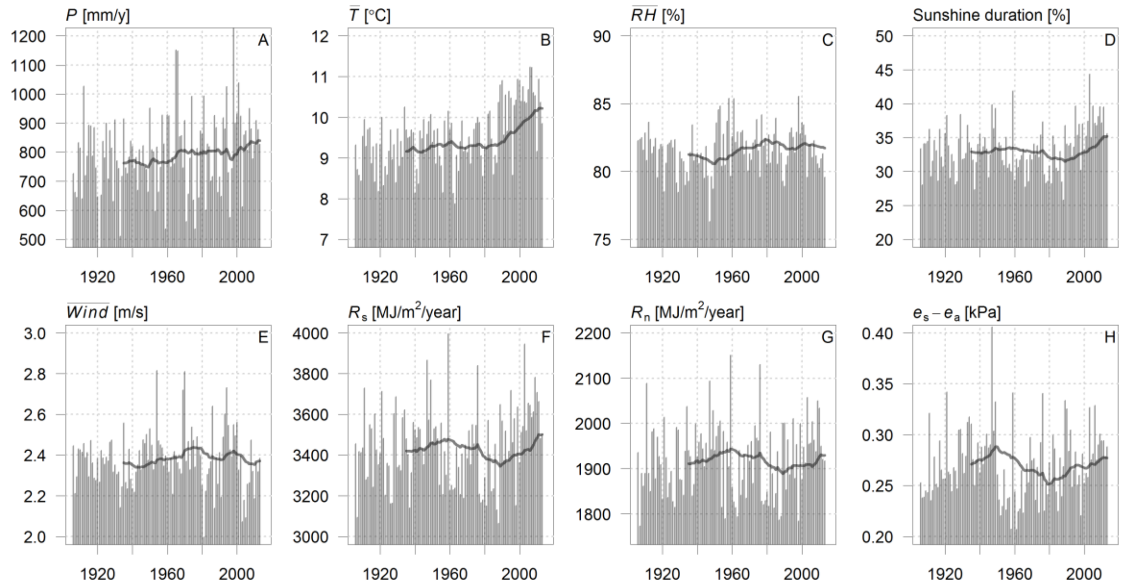

Figure 3: Annual and 30-year moving average variables for De Bilt meteorological station. A: precipitation, B: mean temperature, C: mean relative humidity, D: sunshine duration, E: mean wind speed, F: global radiation, G: net radiation, H: vapour pressure deficit. A-E are observations, F-H are calculated following Allen et al. (1998).

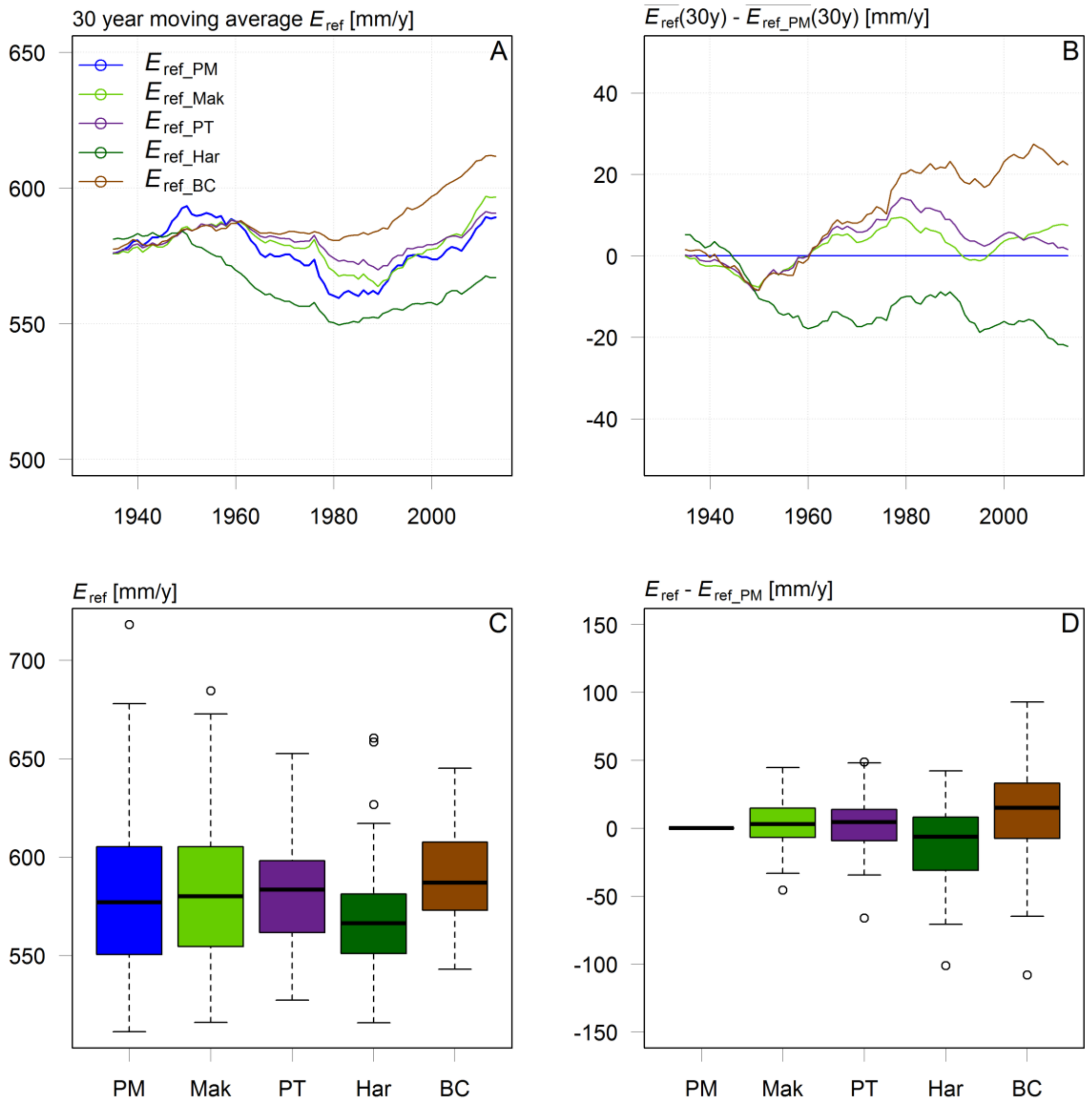

Figure 4: E_{ref} values for five methods for the period 1906-2013. Each empirical method calibrated on daily E_{ref_PM} for the period 1906-1935. A: 30 year moving average E_{ref}, B: deviation of 30 year moving average E_{ref} from E_{ref_PM}. C: yearly variability in E_{ref} for each method. D: yearly deviation of each E_{ref} with E_{ref_PM}. The boxplots show the minimum, first quartile, median, third quartile, maximum and outliers of the annual data.

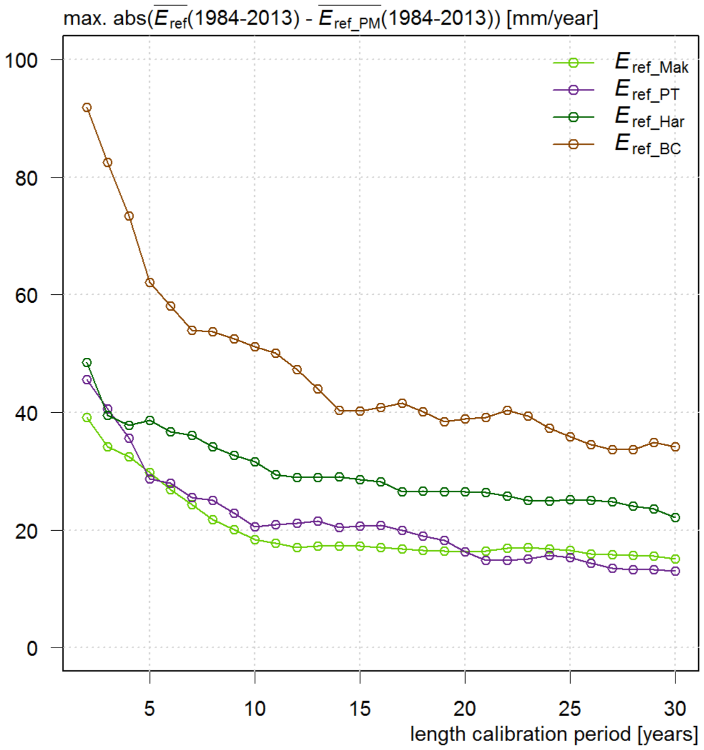

Figure 5: Maximum absolute deviation in 30-year average E_{ref} from 30-year average E_{ref_PM} for the period 1984-2013, as a function of the length of the calibration period.

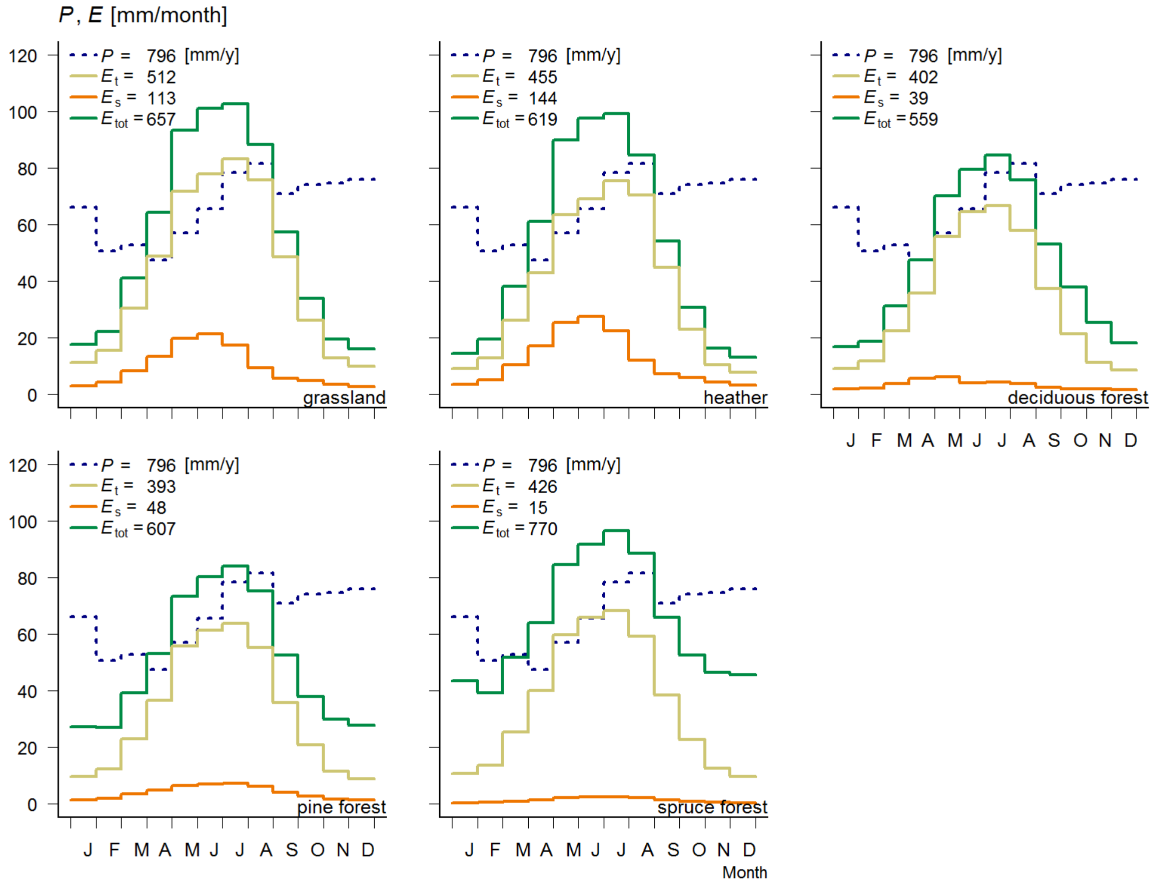

Figure 6: Illustration of synthetic ‘observed’ potential evaporation components simulated with SWAP. The lines give monthly means over the period 1906-2013. E_t and E_s hold for a vegetation stand with a dry canopy only; E_{tot} includes interception.

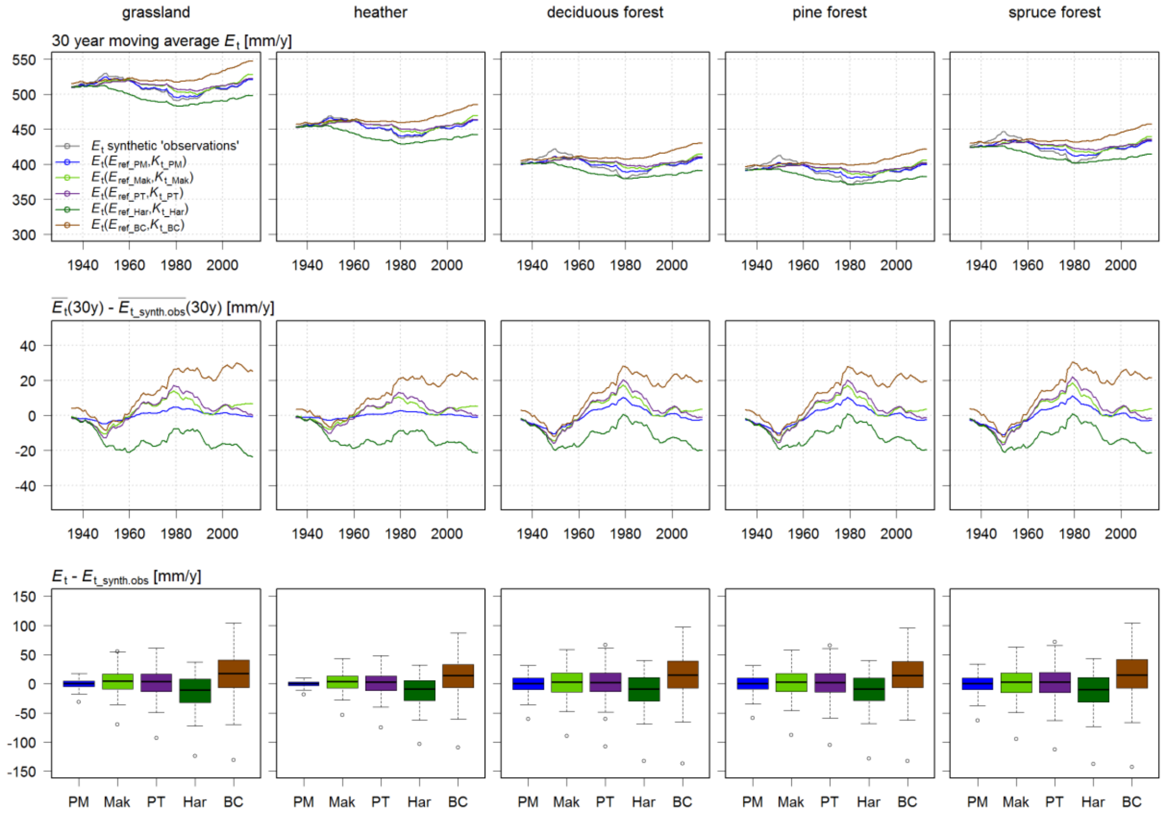

Figure 7: E_t calculated for each vegetation class using each E_{ref} method and matching K_t calibrated on the 30-year period 1906-1935. Presented are 30-year moving averages in mm (top), and deviations with the synthetic E_t for both the 30 year moving averages (center) and annual values (bottom).

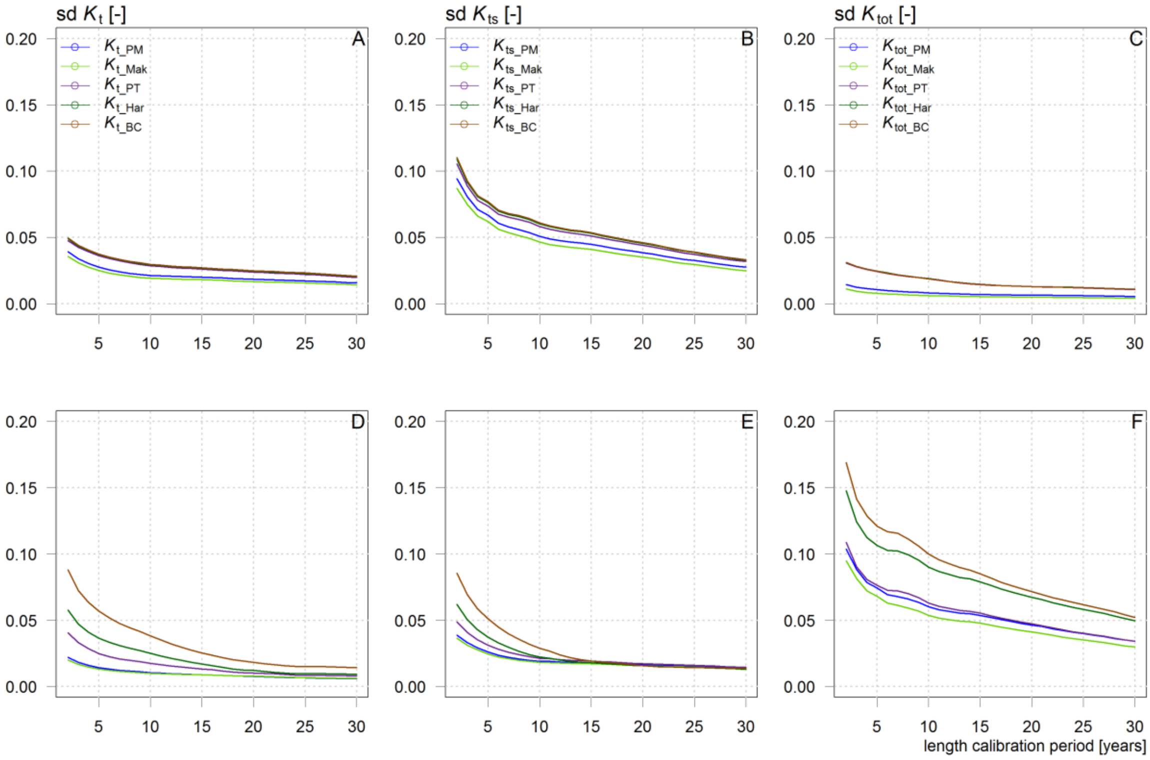

Figure 8: Standard deviation (sd) for K_t (A,D), K_{ts} (B,E) and K_{tot} (C,F) normalized to their mean values, for heather (A-C) and spruce (D-F), as function of the length of the calibration period. K-values are derived for each E_{ref} method; K and E_{ref} are calibrated on the same periods.

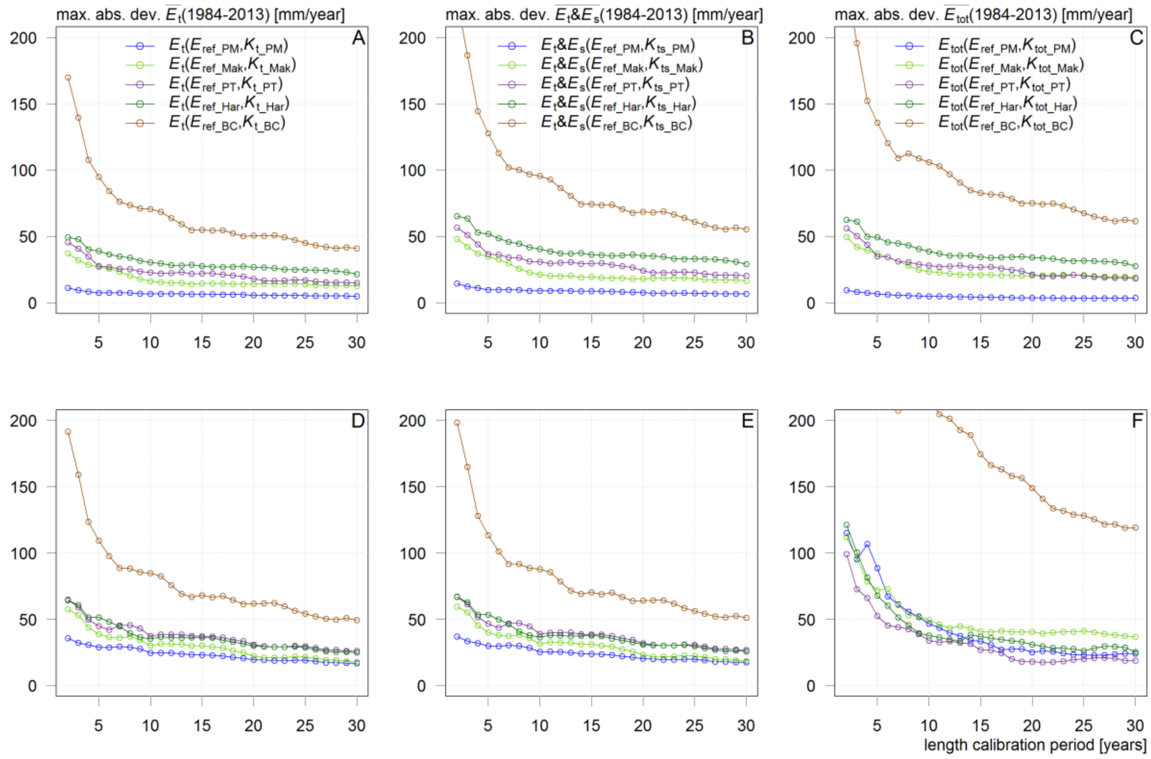

Figure 9: Maximum absolute deviation with synthetic ‘observations’ in mean E_t (A,D), $E_t\&E_s$ (B,E) and E_{tot} (C,F) for the period 1984-2013, for heather (A-C) and spruce (D-F), obtained by the two-step approach, as function of the length of the calibration period. Presented as in Fig. 5, though using E_{ref} and crop factors (K_t, K_{ts} and K_{tot}) to derive E_t, $E_t\&E_s$ and E_{tot}.

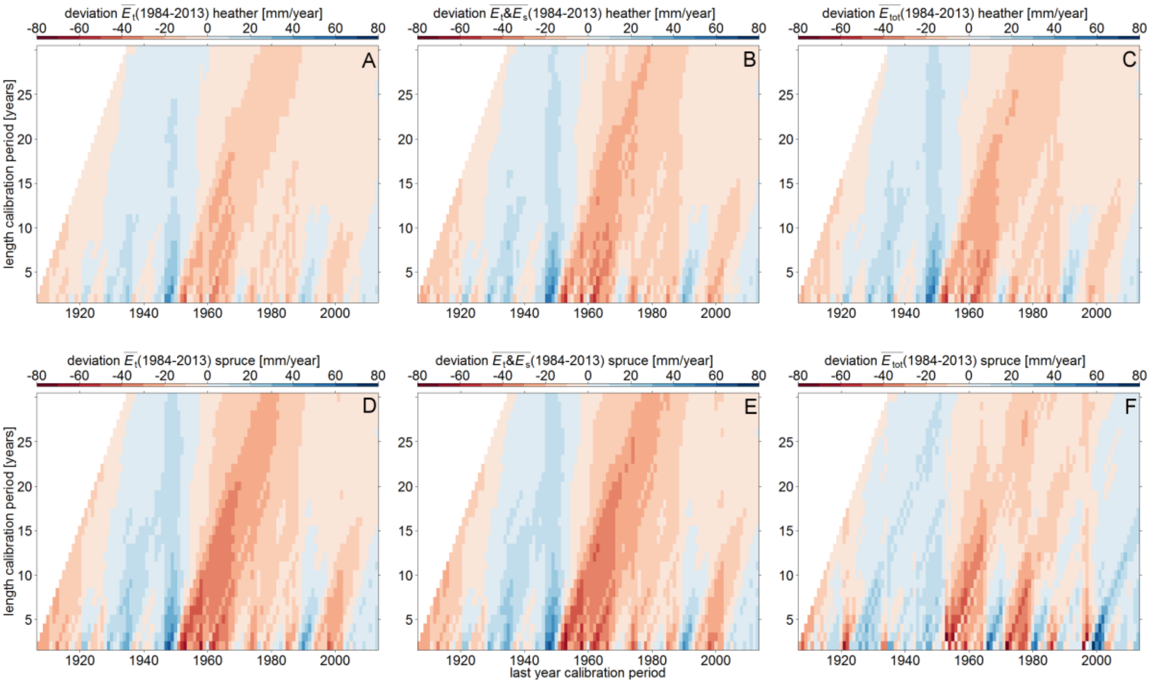

Figure 10: Deviations with synthetic ‘observations’ in E_t (left), E_t&E_s (centre) and E_{tot} (right) for the last 30 year period (i.e. 1984-2013), due to different reference years and lengths of calibration periods for both E_{ref} and K_t, K_{ts} and K_{tot}. Results for PT and heather (top) and spruce (bottom).

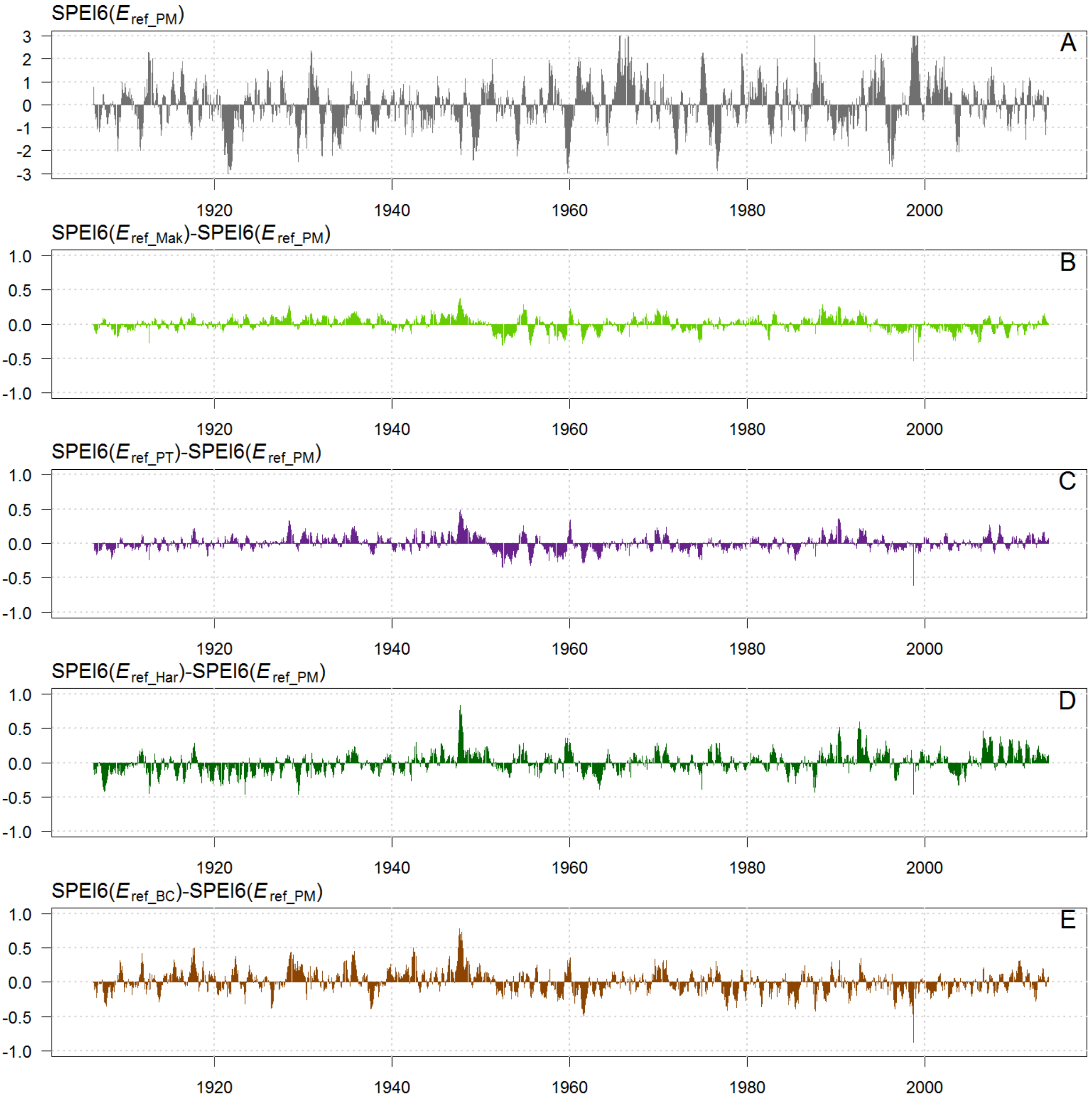

Figure 11: SPEI6 time series with E_{ref} based on PM (A). The subsequent figures show differences in SPEI with E_{ref} based on Mak (B), PT (C), Har (D) and BC (E), calibrated on the period 1906-1935.