Reduction of predictive uncertainty in estimating irrigation water requirement through multi-model ensembles and ensemble averaging

S. Multsch[1], J.-F. Exbrayat[2,3], M. Kirby[4], N. R. Viney[4], H.-G. Frede[1] and L. Breuer[1]

[1]{Institute for Landscape Ecology and Resources Management (ILR), Research Centre for BioSystems, Land Use and Nutrition (IFZ), Justus Liebig University Giessen, Heinrich-Buff-Ring 26, 35390 Giessen, Germany}

[2]{School of GeoSciences and National Centre for Earth Observation, University of Edinburgh, Edinburgh, UK}

[3]{Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, New South Wales, Australia}

[4]{CSIRO Land and Water, GPO Box 1666, Canberra, ACT 2601, Australia}

Correspondence to: S. Multsch (Sebastian.Multsch@umwelt.uni-giessen.de)

Abstract

Irrigation agriculture plays an increasingly important role in food supply. Many evapotranspiration models are used today to estimate the water demand for irrigation. They consider different stages of crop growth by empirical crop coefficients to adapt evapotranspiration throughout the vegetation period. We investigate the importance of the model structural versus model parametric uncertainty for irrigation simulations by considering six evapotranspiration models and five crop coefficient sets to estimate irrigation water requirements

for growing wheat in the Murray-Darling Basin, Australia. The study is carried out using the spatial decision support system SPARE:WATER. We find that structural model uncertainty among reference ET is far more important than model parametric uncertainty introduced by crop coefficients. These crop coefficients are used to estimate irrigation water requirement following the single crop coefficient approach. Using the Reliability Ensemble Averaging (REA) technique, we are able to reduce the overall predictive model uncertainty by more than 10%. The exceedance probability curve of irrigation water requirements shows that a certain threshold, e.g. an irrigation water limit due to water right of 400mm, would be less frequently exceeded in case of the REA ensemble average (45%) in comparison to the equally weighted ensemble average (66%). We conclude that multi-model ensemble predictions and sophisticated model averaging techniques are helpful in predicting irrigation demand and provide relevant information for decision making.

1 Introduction

1.1 Predicting crop water needs

Globally, the proportion of fresh water consumption by agriculture from rainfall as well as surface and groundwater resources is large (9,087 km^3 y^{r-1}) (Hoekstra and Mekonnen, 2012). It is projected that water demand is increasing in the future, in particular by irrigation agriculture, in order to support the increasing world population with food (Foley et al., 2011; De Fraiture and Wichelns, 2010; Hanjra and Qureshi, 2010; Wada and Bierkens, 2014). Therefore, strategies based on improved irrigation methods and local adaptions of management practices are likely to be implemented to anticipate this trend. Such strategies are often developed using decision support systems that are informed by mathematical models. For example, irrigation management has been optimized by modelling and measurements for crops grown in Central Asia (Pereira et al., 2009) or for irrigated cotton in the High-Plains region of Texas (Howell et al., 2004). Others have investigated water use efficiency (Wang et al., 2001) or crop water productivity (Liu et al., 2007) by modelling experiments for irrigated crops grown in China.

All these models depend on the calculation of evapotranspiration (ET) which represents the evaporation from a surface and transpiration from plants. In the case of agricultural crops, ET is equal to the crop water needed for crop growth and yield production. Globally, evapotranspiration represents about two thirds of the total rainfall on land, while evapotranspiration from crops amounts for about 8% (Oki and Kanae, 2006), and is insofar the most important term of the water balance. The basic concept for deriving crop water needs of irrigated crops has been initially reported by Jensen (1968) and is proposed by Allen et al. (1998) as the single crop coefficient concept. The crop specific evapotranspiration (ET_c) is derived from reference evapotranspiration (ET_o) and a crop specific coefficient (K_c):

$$ET_c = ET_o \cdot K_c \qquad (1)$$

with ET_o given in [mm] and dimensionless K_c. ET_o can be calculated by standardise potential evapotranspiration (PET) to a short (grass) or tall (alfalfa) reference crop. In the case of the Penman-Monteith equation (Monteith, 1965; Penman, 1948) standardized fixed values for albedo (0.23), plant height (0.12 cm) and surface resistance (70 m s^{-1}) are assumed (Allen et al., 1998; Jensen et al., 1990). K_c is commonly calculated on the basis of field experiments (e.g. Ko et al., 2009; da Silva et al., 2013) and varies with the crop development.

Such an approach is part of many irrigation management models, including Cropwat (Smith, 1992), ISAREG (Pereira et al., 2009), ISM (George et al., 2000) or global crop water models (Siebert and Döll, 2010). Moreover, the single crop coefficient concept is the basis for the simulation of crop water needs in many studies. For example, Lathuillière et al. (2012) have derived water use by terrestrial ecosystems and have shown that actual ET declines over a 10 year period by about 25% in response to deforestation and replacement by agriculture in Brazil. They showed that irrigation water requirement (IRR) is relevant for terrestrial water fluxes and a reliable estimation is crucial for the closure of the water cycle. In another study future climate impacts on groundwater in agriculture areas have been investigated (Toews and Allen, 2009). They showed that larger return flows to the groundwater can be related to increased IRR under warmer temperatures and longer vegetation periods. Moreover, the crop coefficient concept is also the basis for the water footprint (volume of water consumed or polluted to produce one unit of biomass) assessment of crops (Mekonnen and Hoekstra, 2011) and has been used to determine water requirements and the water footprint of the agriculture sector in Saudi Arabia (Multsch et al., 2013). In almost all studies, researcher use only one ET_o method with a single, often spatially independent K_c set. As a result, some scientists ask to use locally adapted K_c sets at least (Ko et al., 2009; da Silva et al., 2013). For this reason, the investigation of predictive uncertainty of IRR is needed, in particular in the frame of large scale assessments.

1.2 Sources of predictive uncertainty

Major sources of uncertainties should be considered in the study design, quantified throughout the modelling process (Refsgaard et al., 2007) and communicated as part of the results to the end users. Uncertainties related to large scale estimations of the IRR have only rarely been analysed. For example, Siebert and Döll (2010) have studied the uncertainty in predicting green (rainfall consumed by crops) and blue (consumed surface and groundwater by crops in terms of irrigation) water consumption by using different ET_o equations on a global scale. They observed a significant difference of blue water consumption, i.e. required irrigation, and only a small change in green water consumption between model runs while using two classical ET_o equations. More recently, Sheffield et al. (2012) pointed out that using a more up-to-date parameterization of PET to calculate drought indices led to different conclusions on drought occurrence globally.

Generally, model predictive uncertainty can be lead back to four sources, input uncertainty, output uncertainty, structural uncertainty and parametric uncertainty (Renard et al., 2010). The last two, structural and parametric uncertainty, are addressed in this study with a focus on the prediction of IRR. As part of the parametric uncertainty, the parameterization of equations to quantify natural or anthropogenic processes has received considerable interest, particularly in conceptual rainfall-runoff modelling (Beven, 2006; Vrugt et al., 2009). In case of modelling crop water needs according to Eq. 1, K_c is an important model parameter. K_c values for a large number of crops are provided by the FAO56 irrigation guidelines (Allen et al., 1998) which are commonly used for irrigation planning. However, it has been highlighted that an adjustment to the global K_c is needed if the simulations are used for irrigation planning on a local to regional scale (Ko et al., 2009; da Silva et al., 2013). Nevertheless, it is still unclear whether a local adaption of K_c leads to a better model performance. For this reason, we quantify the parametric uncertainty of model parameterisation with different K_c sets.

The model structure also introduces uncertainties, as any model remains a simplification of the real world. In the context of modelling water resources, all hydrological and crop growth models rely on the estimation of ET. According to equation 1, ET_o is required to estimate crop specific evapotranspiration. ET_o equations are often divided into categories according to the input data (Bormann, 2011; Tabari et al., 2013): temperature based equations such as Hargreaves-Samani (HS) equation (Hargreaves and Samani, 1985), radiation based equations such as Priestley-Taylor (PT) (Priestley and Taylor, 1972) or combined equations such as the FAO56 Penman-Monteith (PM56) equation (Allen et al., 1998), that further takes wind speed into account. Nevertheless, in many cases it was shown that the variability among ET_o methods is large (Fisher et al., 2011; Kite and Droogers, 2000). Because most water resources models rely on some calculation of ET_o, we see it as a crucial source of structural uncertainty that is rarely considered.

1.3 Reduction of predictive uncertainty by ensemble modelling

Ensembles of model predictions can be developed by different sets of model parameterization (single-model ensemble) and model structures (multi-model ensemble). The weighting of model ensembles according to their fit to observational data has become of interest to reduce the uncertainty and to derive a more robust predictions and projections. Giorgi and Mearns (2002) have introduced the reliability ensemble averaging technique (REA) in climate research. Basically, different models are weighted according to their performance in representing measured data and according to the distance of individual models to the ensemble average prediction to quantify the convergence of different models. This approach has been applied more recently for predicting catchment nitrogen fluxes (Exbrayat et al., 2013) and calculating water balances and land use interaction (Huisman et al., 2009).

In a first step, we analyse the relative contributions of the structural and parametric model uncertainty in hind casts of IRR of wheat across the Murray-Darling-Basin (MDB), Australia. Simulations are calculated using the spatial decision support system SPARE:WATER (Multsch et al., 2013). In a second step, we apply the REA methodology to reduce the predictive uncertainty of IRR. The general procedure is as follows:

- The applicability of six different ET_o methods is evaluated by using available measured class-A-pan evaporation measurements of 34 stations in the MDB over a 21years time period;
- 30 different model realisations are setup in a multi-model ensemble by combining various ET_o equations (n=6) and crop coefficient data sets (n=5);
- IRR is calculated by forcing the multi-model ensemble with climate time series of 21 years (monthly data) for 3,969 sites (each 1 km² x 1km²) in the MDB where irrigated wheat has been grown according to the land use allocation in 2000;
- The 30 model realisations are weighted according to their performance in representing measured data and their distance to the ensemble average.

By doing so, we quantify structural (ET_o method) and parametric (K_c set) uncertainty and apply REA to provide a robust estimate of IRR and the confidence interval around it. The underlying research question is how can we derive better predictions by using an ensemble of well-known ET_o methods as well as K_c sets and which are the likely causes of predictive uncertainty in IRR estimations. Finally, we show a procedure to reduce predictive uncertainty of IRR.

2 Methods and data

2.1 Study site and data

The MDB covers about 1 million km² of south-east Australia (Fig. 1). Irrigation agriculture in the MDB sums up to 17,600 km², which is equal to 65% of the total irrigation agriculture in Australia. Total water withdrawal for irrigation in 2006 amounted to 7.36 km³ yr^{-1} (ABS, 2006). Wheat is the second most important crop grown in MDB after grazing pastures, covering 3,969 km² in 2006 and was therefore selected for this case study for which IRR and its underlying uncertainty was calculated. The cropping areas have been taken from a land use map from 2006 (ABARES, 2010) with a spatial resolution of 0.01° x 0.01° (~1 km x 1 km). We assume a fixed land use distribution over time in our model study to clearly target the uncertainty in ET_o method and crop coefficients. Climate data for 1986-2006 were taken from the SILO Data Drill of the

Queensland Department of Natural resources and Water (https://longpaddock.qld.gov.au/silo/ (Jeffrey et al., 2001)) with a spatial resolution of 0.05° x 0.05° (~5 km x 5 km). We used the same weather dataset over all 3,969 1 x 1 km land grid cells overlapped by a 5 x 5 km grid cell in the weather data. The model was forced with monthly data. For validation, we compared simulated ET_o to measured class-A pan data from 34 stations throughout the MDB. The class-A pan data were obtained from Patched Point Dataset of the Queensland Department of Science, Information Technology, Innovation and the Arts, (http://www.longpaddock.qld.gov.au/silo/ppd/). Measured data have been adjusted with monthly pan-coefficients according to McMahon et al. (2013) to represent evaporation from open surface water. For stations where no pan-coefficient was available we used the one from the nearest station.

2.2 Simulation of irrigation requirement with SPARE:WATER

SPARE:WATER (Multsch et al., 2013) is a spatial decision support system for the calculation of crop specific water requirements and water footprints from local to regional scale. Input parameter for the simulation are climate data, irrigation management (irrigation water quality, irrigation efficiency, irrigation method), a digital elevation model and crop characteristics such as maximum crop height and length of growing season as well as sowing and planting date. In a first step, the water requirement of growing a crop is simulated for each grid cell according to the spatial resolution of the input data. In a second step, the water footprint for spatial entities such as administrative boundaries or catchments is calculated considering statistical data on crop yield and harvest area. Water footprints for geographic entities are given as volume of water consumed per year (e.g. $km^3\ yr^{-1}$) and water footprints for specific crops as volumes of water consumed per biomass ($m^3\ t^{-1}$).

In this study the calculation of the IRR is calculated as the difference between ET_c and effective rainfall (P_{eff}). The latter one is estimated from the difference of surface run-off (RO) and precipitation (P). RO is derived as a fixed fraction of 20% of total P. The fixed fraction of runoff is adapted from the default setting of the FAO Cropwat model (Smith, 1992). On this basis, IRR is calculated according to Eq. 2:

$$IRR = \max\left(ET_c - P_{eff}, 0\right) \qquad (2)$$

with IRR, ET_c and P_{eff} given in [mm]. ET_c is calculated based on the single crop coefficient approach initially proposed by Jensen (1968) and recommended by Allen et al. (1998) according to Eq. 1. The input parameters for this method are the length of four individual stages (initial season, growth season, mid-season and late season) during the growing season and three related crop coefficients (K_c). These define the ratio between ET_o and ET_c for each part of the growing season. We have considered five different K_c data sets (Table 1). The most common dataset has been proposed from the FAO56 Irrigation and Drainage Guidelines (Allen et al 1998). This approach has been applied for calculating crop water footprints (Mekonnen and Hoekstra, 2011) and is part of the widely used Cropwat model (Smith 1992). It has been discussed that locally adapted K_c sets are superior in simulating site-specific crop water requirement than global ones (Ko et al., 2009; da Silva et al., 2013). Thus, further data sets have been collected from various sources which represent site-specific relationships between ET_o and ET_c for areas in the MDB.

ET_o has been calculated with six different methods (Table 2). Two of them are classified as combined methods (PM56, PPET), three are radiation-based methods (PT, TURC, APET) and one is a temperature based method (HS). All of them are commonly applied function, e.g. PM56 and HS are included in Cropwat (Smith, 1992) and Aquacrop (Steduto et al., 2009), two models to quantify crop water and IRR, widely used and promoted by the FAO. The cropping system model EPIC (Williams, 1989) additionally allows the use of the PT equation, while the global vegetation model LPJmL (Fader et al., 2010) and the global water model WaterGap (Döll et al., 2003) are restricted to PT. APET and PPET have been particularly tested for the utilisation under Australian weather conditions in several (Chiew et al., 2002; Chiew and Leahy, 2003; Donohue et al., 2010).

2.3 Reliability Ensemble averaging

We used two types of ensemble averaging techniques, which differ in the weighing technique. We calculated an equally weighted average of all 30 model realisations (6 ET_o methods x 5 K_c datasets) for every grid cell which sum up to 3,969 cells (1 x 1 km) in the MDB where irrigated wheat is grown according to the land use allocation in 2006. However, this method does not consider the capability of its ensemble members to predict a target value nor does it value the agreement of model predictions amongst each other. Therefore, we apply the REA technique that was initially proposed by Giorgi and Mearns (2002) to reduce uncertainties in climate change projections. Moreover, it was used in impact studies targeting land use change impacts on hydrology (Huisman et al., 2009) and water quality scenario projections (Exbrayat et al., 2013).

The strength of the REA method is that it considers both the quality of a model prediction (performance) and its position within an ensemble of prediction (convergence). The aim is to provide a best estimate of predictions and a robust assessment of the confidence interval around it. The REA weighting scheme estimates two factors, model performance (R_B) and model convergence (R_D). R_B represents the capability of each ensemble member to represent real world data by its bias B. R_D is a measure of the distance D of a single model to the equally weighted ensemble average. Both are limited by the natural background variability (ε). The combined effect known as reliability factor (R) is derived as:

$$R = \overbrace{\left[\frac{\varepsilon}{abs(B)}\right]}^{R_B} * \overbrace{\left[\frac{\varepsilon}{abs(D)}\right]}^{R_D} \qquad (3)$$

In this study, ε is calculated from measured class-A pan evaporation for 34 climate stations in the study region for the time period from 1986 to 2006. The class-A pan data has been adjusted with monthly pan coefficients for climate stations in Australia (McMahon et al., 2013). We calculated the annual mean evaporation [mm] for each year and each station and used the 50% confidence interval (difference between the 25% and 75% percentile) of 224 mm to define ε. The consideration of the difference between upper and lower percentiles has been recommended by Giorgi and Mearns (2002). Model performance is measured by the RMSE between measured (class-A pan) and predicted ET_o for each model (i).

The convergence criterion R_D is calculated in an iterative procedure. The difference between the average IRR of each ensemble member *i* and the ensemble average is calculated. Under the consideration of the natural background variability ε a first guess of R_D (for each ensemble member) is predicted as well as a first guess of the REA average. This procedure is repeated by considering the newly derived REA average until the ensemble convergence, so that the difference between ensemble members and the REA average cannot be reduced by additional iterations (see Giorgi and Mearns (2002) for a complete methodological description). The error of the equally weighted ensemble average is described by the RMSE between IRR_i predicted by model *i* (with n=30 models) and the equally weighted ensemble average irrigation water requirement ($\overline{IRR}$). The error of the reliability ensemble average ($RMSE_{REA}$) is derived from the reliability factor of each model (R_i), the irrigation water requirement predicted by model *i* (IRR_i) and the REA weighted ensemble average ($\overline{IRR}_{REA}$). The RMSE represents an approximate 60-70% confidence interval under the assumption that the amount of irrigation is distributed somewhere between normal and uniform.

3 Results

3.1 Validation of ET_o methods

We applied six ET_o equations to 34 sites in the MDB for which measured class-A pan evaporation data were available from 1986 to 2006 (Fig. 2). Class-A pan data represent the evaporation from an open water surface and integrate all climate factors driving evaporation such as radiation, wind speed, humidity and temperature. Pan evaporation differs from evaporation from a cropped surface through a different albedo, heat storage and humidity above the surface. For this reason, the class-A pan data have been adjusted with monthly pan coefficients (McMahon et al., 2013) to better compare them with ET_o simulations of open surface waters. On an annual average, class-A pan evaporation of 1,558 mm yr^{-1} were reduced by 9% to 1,422 mm yr^{-1} across all stations.

The median daily ET_o for APET is 3.6 mm d^{-1}, PM56 3.9 mm d^{-1}, HS 3.8 mm d^{-1}, PPET5.2 mm d^{-1}, PT 6.4 mm d^{-1} and TURC 3.4 mm d^{-1}. According to the root-mean-squarederror (RMSE) PM56 gave the most reliable results. The median of ET_o for APET, PM56 and HS are close to the median of the measured evaporation rate of 3.7 mm d^{-1}. Apart from PT and PPET, the other methods underestimate ET_o, especially where class-A pan data are larger than 6 mm d^{-1}. The relationship between measured and simulated ET_o is linear as shown by the coefficients of determination r^2 ranging from 77% (PT) to 88% (PPET).

The simulated ET_o is normally distributed if a single station and one year is tested (Shapiro test for normality: alpha>0.1 for each year and station). The difference between the 34 stations is up to two times larger than the inter-annual difference in the 21 years period. Thus, spatial variability is larger than temporal variability in the MDB. The intra-annual variability shows a different picture. The median ET_o in the summer months is up to four times larger than the ET_o during winter months for all ET_o methods, except PPET and PT with a six times larger ET_o in summer than in winter months.

Four of the six methods simulate the measured data with a high r^2 and a low RMSE. The difference between the methods itself is large, in particular through the high ET_o estimates by PT and PPET. Thus, the structural uncertainty through the ET_o method is substantial and needs to be considered for the prediction of IRR which is addressed in the next chapters.

3.2 Irrigation water requirement and its variability

The IRR of wheat has been simulated using an ensemble of thirty model realisations for each of the 3,969 1 km × 1 km irrigated cells in the MDB for 21 years. Average values of IRR for all model realisations are shown in Table 3. In most cases, the largest estimates are given by the combinations of the Kc set Hughes with the ET_o method PT. These are almost 2.5 times higher than the lowest average IRR calculated by the combination of TURC with the Kc set Harris. It is obvious that changing ET_o method results in a larger variation of calculated IRR than using a different Kc set. Hence, the average IRR give a first idea about variability due to model structures and parameters.

Over a large watershed such as the MDB local differences in IRR may be large while catchment wide water management plans define thresholds for water withdrawal, for example due to water rights or water resources protection measures. A given threshold may require heterogeneous local adaptations of irrigation management and a change in cropping patterns. Figure 3 shows the probability that a certain amount of IRR is exceeded in the MDB on average over the 21 year period. It illustrates the range of IRR predicted by the ensemble of all 30 model realisations for each grid cell. Two groups can be identified that are separated by ET_o methods. The first group is composed of PPET and PT calculations. In this case, IRR is up to twice as high as compared to predictions by other models. The second group is formed by APET, HS, PM56 and TURC with substantially lower calculations of less than 500 mm in most cases. We note that the parametric uncertainty is almost negligible compared to the uncertainty introduced by the various ET_o methods.

3.3 Ensemble averaging, uncertainty and weighting

Ensemble predictions have become an important tool to account for different model structures and parameters (Exbrayat et al., 2013; Huisman et al., 2009; Wada et al., 2013). The consideration of ensembles is especially helpful to increase our confidence in simulations when no validation data are at hand, such as projections of Earth's future climate under specified emission scenarios. Here we apply the concept of ensemble prediction to simulations of IRR. Two different ensemble averages, expressed as the exceedance probability of the IRR of wheat are shown in Fig. 4. The first one represents the equally weighted average of irrigation ($\overline{IRR}$, black line). The second one represents a weighted average using the reliability ensemble averaging ($\overline{IRR}_{REA}$, red line, see methods description) that weights predictions based on their performance and agreement with other ensemble members. This prevents dismissing some model structure, a process that can be rather subjective. Also, even an overall poorly performing model can contribute to the optimal information extracted from the ensembles (Viney et al., 2009), or may outperform better performing models once boundary conditions are changed (Exbrayat et al., 2013).

We use the inverse of the cumulative daily RMSE (Fig. 2) of the ET_o methods during the growing season to calculate the criterion R_B (RMSE 154 mm for APET, 123 mm for PM56, 142 mm HS, 232 mm PPET, 373 mm PT, 166 mm TURC). The convergence criterion R_D was calculated based on the difference of the predicted irrigation given by a single ensemble member and the equally weighted ensemble average (see Methods description). Overall, the PT model combinations have the lowest reliability factors of between 0.51 and 0.6 followed by PPET with 0.96, a result driven by the poorer performance of these methods to simulate pan-evaporation (Fig. 2), and the outlying positions of simulations using PT and PPET (Fig. 3). All other models are weighted similarly, a result in accordance with the similar performance and simulated values exhibited by these methods (see Table 4 for details).

The application of the reliability factor leads to a decrease of the calculated total IRR in each grid cell as well as to a decrease of its overall uncertainty (Fig. 4). The uncertainty range is given by the ensemble average plus/minus the RMSE in each grid cell, assuming that modelling errors are normally distributed.

Exceedance probability curves might support defining thresholds in irrigation planning with consequences for decision makers through, for example, the adaptation of improved irrigation practice (e.g. from full to deficit irrigation, installation of advanced irrigation techniques) or the purchase of additional water rights. For example, a limit of available irrigation water of 400 mm per growing season will be exceeded less frequently in the MDB if the REA average IRR is considered (45%) in comparison to the equally weighted average (66%).

The spatial distribution of the equally weighted and the REA weighted ensemble averages are shown in Fig. 5a and b. The equally weighted average of IRR ranges between 124 and 691 mm with an average across the MDB of 424 mm (Fig. 5a). Thus, spatial variability is large and western and northern areas require five to six times more irrigation than in the south-east. The REA derived average IRR ranges between 104 mm and 663 mm across the river basin (Fig. 5b) with an average of 405 mm. Depending on the location this value is up to 18% lower as compared to simulations based on the equally weighted average (Fig. 5c). Also, the uncertainty range decreases as consequence of the REA method by about 10 % across the MDB with maximum values of around 26% when comparing equally and REA weighted RMSE (Fig. 5d-f). The largest change in uncertainty can be found in the south-east of the MDB and also in areas towards the east (Fig. 5f). Thus, REA not only leads to a decrease of predicted IRR but also to a reduction of its uncertainty. The uncertainty is reduced because the REA is drawn toward the group of the better ET_o methods that also agree well between themselves.

4 Discussion and conclusions

The simulation of IRR strongly varies amongst ET_o methods. Bormann (2011) recommended that the selection of the ET_o method should be based on the validation of ET_o with real world observations rather than only on the availability of climate input data. This is due to the general large variability among ET_o methods, which was also revealed in a study where PT was set as a benchmark model and the RMSE between ET_o methods was analysed (McMahon et al., 2013). Likewise, the influence of a single ET_o method on the prediction of crop yields was also reported for an agriculture site in Europe (Balkovič et al., 2013) where ET_o estimates by PT were 40% higher and those by Penman-Monteith 10% lower in comparison to HS. We also found a large variability among ET_o methods in our study. However, similar ranges across Australia for ETo have been reported by others (Chiew et al., 2002) for APET, PPET and PM56 as well as lower values for PT. Lascano *et al* (2010) as well as Lascano and Van Bavel (2007) have shown that methods to calculate ET_o based on combination methods, i.e., Penman-Monteith, tend to underestimate ET by as much as 25%, especially in dry climates.

Bormann (2011) further recommended that the reliability of ET_o equations should be tested in a spatial context, especially if applied on large scale. For various regions across Australia, a large range of mean annual ET_o between 1,700 mm (PT) and 3,670 mm (PPET) was reported (Donohue et al., 2010). To investigate the spatial heterogeneity within the MDB we analysed results of the 34 class-A pan stations. Overall, the performance of four of the ET_o methods was good with RMSEs around 1 mm day^{-1}, except for three stations in the north. PPET performed less well with RSME increasing to 2 mm day^{-1} while the PT value ranged up to 4 mm day^{-1}. However, we found no consistent spatial pattern. We are aware that the utilization of class-A pan data comes along with uncertainties. We did not assume that the data are error-free, but for the application of REA, a comparison of model simulations and observations is needed to calculate the model performance criterion. We could have treated PM56 as being an “observation” in terms of a benchmark model. However, we think that a more independent test is more appropriate in the sense of REA and therefore decided to use those observations that are at hand: class-A pan observations. To account for the difference of class-A pan evaporation and reference crop ET, we used a commonly applied correction factor (pan-coefficients according to McMahon et al. (2013)) to derive crop ET from class-A pan measurements. Most often, ET estimates are not compared to any measurements at all, leaving modelers with no information on how good their model application is. We therefore think that a comparison to class-A pan is for sure not perfect, but better than not testing at all.

ET_o estimates using the PM56 method revealed the best performance criteria in our study. PM56 considers the most meteorological input parameters thereby possibly best representing the altering dry and wet conditions across the MDB over the year. The better performance of physically based equations in comparison to more empirical approaches for the simulation of ET_o has also been reported by others (Donohue et al., 2010). PT performed least well in our study and resulted in up to two times larger estimates than other ET_o methods. This is somewhat contrasting with other studies (Chiew et al., 2002; Donohue et al., 2010) where PT gave lower ET_o values in comparison to methods such as APET and PPET. One reason is that Donohue et al. (2010) have considered the actual albedo from remotely sensed vegetation cover (Donohue et al., 2008) for the estimation of the net incoming solar radiation. In our calculations, an albedo of a reference crop 0.23 (short crop, i.e. grass) has been considered according to the guidelines for ET_o from Allen et al. (1998). Another likely reason for this observation is that the PT equation is based on the Penman-Monteith equation in which the aerodynamic term is replaced by a constant (alpha) which is commonly set to 1.26 under Australian climatic conditions (Chiew and Leahy, 2003) and which we also applied. The consideration of region-specific alpha for the MDB could have increased the performance of PT in our study. The HS equation is commonly applied in situations where meteorological data are scarce, because the equation depends on more readily available temperature and extra-terrestrial radiation derived from latitude and day of the year. A reason for its good performance in our study could be that the semi-arid climate in most of the MDB is favourable for the HS equation, which is supported by Tabari (2010) who conclude that HS is a good candidate model for warm humid and semi-arid sites, but fails under cold humid climates. However, the poor response of HS to changing climatic boundary conditions has also been criticized in a study on global drought simulations (Sheffield et al., 2012).

We combined the six ET_o methods with five K_c sets to address stochastic parametric uncertainty for irrigated wheat in the MDB. We show that the ET_o method uncertainty range exceeded the uncertainty range of Kc sets. Thus, the Kc sets have a minor influence on predicted IRR. At first sight, this seems to be contrasting to others who have stated that adapted, regional Kc sets are required to estimate reliable IRR rates. For instance, da Silva et al. (2013) reported that Kc sets from FAO56 lead to errors in plot scale irrigation planning under tropical conditions. Similar observations were reported for semi-arid conditions in the Texas High Plains region (Ko et al., 2009), highlighting the importance of regionally based Kc sets. While regional adaptation of Kc might be important at smaller scales, e.g. on the farm level, we conclude that large scale applications do not necessarily need to focus on this potential contribution of uncertainty. Rather, effort should be put into finding appropriate ET_o methods, or even better, utilize ensemble predictions to cover a more realistic range of predictions. Our study confirms this latter recommendation, as we could not identify a single best ET_o method for the MDB. Especially in cases where no data for a direct evaluation of model results are available the application of model ensembles gives insight to the predictive uncertainty, e.g., being helpful in the development of best management practices (Exbrayat et al., 2013), study of land use (Huisman et al., 2009) or climate change (Exbrayat et al., 2014).

Besides the uncertainty introduced by local to global Kc values the utilisation of the single crop coefficient concept itself comes along with errors, which are not addressed in this study. For example, Lascano (2000) shows how Kc varies as a function of time (50 days) and how it changes when using a daily, 3 and 8-day moving average. Moreover, the temporal resolution of ETo calculation, i.e., hourly vs. daily is an important component and errors associated with the method of irrigation (surface, drip, sprinkler) cannot be neglected, but are beyond the uncertainty calculation of this study. We acknowledge that we do not consider uncertainties in boundary conditions (e.g. land-use management options, climatic variability) although these may be non-negligible. For example, Bocchiola et al. (2013) reported that changes in future precipitation regimes will have the greatest impact on the calculated water footprint (reflecting high ET rates) of maize in Italy and that changes in CO_2 and warming were less important. Conversely, water use was more driven by agricultural management than by regional climatic variation in a water footprint analysed for an irrigation district in China (Sun et al., 2013). Statistical correction of model forcing data (such as bias correction of precipitation) has also been reported to alter ET estimates as shown by Ye et al. (2012) for the Upper Yellow River in China with changes of up to 29% of ET. Beyond that, the forcing data themself introduce additional uncertainties. However, this is not part of this study and it would clearly go beyond the scope of our work presented here. Nevertheless, on the long term more research needs to be put in the investigation of the global predictive uncertainty of models, where all sources of uncertainty are evaluated, i.e. spatial input data uncertainty (e.g. soil and land use information), model forcing data uncertainty (e.g. climate data), parameter uncertainty, and model structure uncertainty. Thus, an even more complete picture of global model uncertainty can only be shown by considering all sorts of predictive uncertainty, including model input data, validation data, and spatial input data in addition to the impact of model structural and parametric uncertainty as presented here.

However, we argue that future management practices or the impact of climate change cannot be reliably evaluated due to the large uncertainty that exists in the ET_o method, the basis of water resources modelling. We partially cope with this problem by applying the REA technique to extract the most relevant information from our simulations. The advantage of REA in decision making has already been shown for other fields of research, such as the development of N reduction scenarios to improve surface water quality (Exbrayat et al., 2013) or estimation of the effect of land use change on water budgets and hydrological fluxes (Huisman et al., 2009). Despite the growing importance of IRR for today’s agriculture (Siebert and Döll, 2010) and the effect on surface (Hoekstra et al., 2012) and groundwater (Wada et al., 2010) resources, few studies have dealt with the predictive uncertainty of this requirement (e.g. Wada et al. (2013)) and how to reduce it.

Acknowledgements

We acknowledge the generous funding of the Deutsche Forschungsgemeinschaft (DFG, grant BR2238/11-1) that allowed a cooperation visit of the first author to CSIRO Land and Water and the ARC Centre of Excellence for Climate System Science in Australia. The research was further supported by a grant from KACST, Saudi-Arabia and the CAWa project (AA7090002).

References

ABARES: Land Use of Australia Version 4 2005/2006, Department of Agriculture Fisheries and Forestry, Australian Bureau of Agricultural and Resource Economics, 2010.

ABS: Water Use on Australian Farms Murray-Darling basin 2005-06, 46180DO012, 2006.

Allen, R. G.: REF-ET user's guide, University of Idaho Kimberly Research Stations, Kimberly, 2003.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, FAO, Rome, 300, 6541, 1998.

Balkovič, J., van der Velde, M., Schmid, E., Skalský, R., Khabarov, N., Obersteiner, M., Stürmer, B. and Xiong, W.: Pan-European crop modelling with EPIC: Implementation, up-scaling and regional crop yield validation, Agr. Syst., 120, 61–75, 2013.

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320(1), 18–36, 2006.

Bocchiola, D., Nana, E. and Soncini, A.: Impact of climate change scenarios on crop yield and water footprint of maize in the Po valley of Italy, Agr. Water Manage., 116, 50–61, 2013.

Bormann, H.: Sensitivity analysis of 18 different potential evapotranspiration models to observed climatic change at German climate stations, Clim. Change, 104(3), 729–753, 2011.

Chiew, F. H. S. and Leahy, C.: Comparison of Evapotranspiration Variables in Evapotranspiration Maps for Australia with Commonly Used Evapotranspiration Variables, Austr. J. Water Resour., 7(1), 1-11, 2003.

Chiew, F., Wang, Q. J., McConachy, F., James, R., Wright, W. and deHoedt, G.: Evapotranspiration maps for Australia, in Hydrology and Water Resources Symposium 2002: the Water Challenge, Balancing the Risks, p. 167, Institution of Engineers, Australia., 2002.

Döll, P., Kaspar, F. and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, J. Hydrol., 270(1), 105–134, 2003.

Donohue, R. J., McVicar, T. R. and Roderick, M. L.: Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate, J. Hydrol., 386(1), 186–197, 2010.

Donohue, R. J., Roderick, M. L. and McVicar, T. R.: Deriving consistent long-term vegetation information from AVHRR reflectance data using a cover-triangle-based framework, Remote Sens. Environ., 112(6), 2938–2949, 2008.

Exbrayat, J.-F., Buytaert, Timbe, E., Windhorst, D. and Breuer, L.: Addressing sources of uncertainty in runoff projections for a data scarce catchment in the Ecuadorian Andes, Clim. Change, 125(2), 221-235, 2014.

Exbrayat, J.-F., Viney, N. R., Frede, H. G. and Breuer, L.: Using multi-model averaging to improve the reliability of catchment scale nitrogen predictions, Geosci. Model Dev., 6, 117–125, 2013.

Fader, M., Rost, S., Müller, C., Bondeau, A. and Gerten, D.: Virtual water content of temperate cereals and maize: Present and potential future patterns, J. Hydrol., 384(3), 218–231, 2010.

Fisher, J. B., Whittaker, R. J. and Malhi, Y.: ET come home: potential evapotranspiration in geographical ecology, Global Ecol. Biogeogr., 20(1), 1–18, 2011.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D. and Zaks, D. P. M.: Solutions for a cultivated planet, Nature, 478(7369), 337–342, doi:10.1038/nature10452, 2011.De Fraiture, C. and Wichelns, D.: Satisfying future water demands for agriculture, Agr. Water Manage., 97(4), 502–511, 2010.

George, B. A., Shende, S. A. and Raghuwanshi, N. S.: Development and testing of an irrigation scheduling model, Agr. Water Manage., 46(2), 121–136, 2000.

Giorgi, F. and Mearns, L. O.: Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging" (REA) method, J. Clim., 15(10), 1141–1158, 2002.

Hanjra, M. A. and Qureshi, M. E.: Global water crisis and future food security in an era of climate change, Food Policy, 35(5), 365–377, doi:10.1016/j.foodpol.2010.05.006, 2010.

Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, Appl. Eng. Agric., 1(2), 96–99, 1985.

Harris, G. A.: Irrigation: water balance scheduling, Queensland Department of Primary Industries and Fisheries, (DPI Note FSO546), 2002.

Hoekstra, A. Y. and Mekonnen, M. M.: The water footprint of humanity, P. Natl. Acad. Sci. USA, 109(9), 3232–3237, 2012.

Hoekstra, A. Y., Mekonnen, M. M., Chapagain, A. K., Mathews, R. E. and Richter, B. D.: Global monthly water scarcity: blue water footprints versus blue water availability, PlOS One, 7(2), e32688, 2012, DOI: 10.1371/journal.pone.0032688

Howell, T. A., Evett, S. R., Tolk, J. A. and Schneider, A. D.: Evapotranspiration of full-, deficit-irrigated, and dryland cotton on the Northern Texas High Plains, J. Irrig. Drain. E., 130(4), 277–285, 2004.

Hughes, J. D.: Southern Irrigation SOILpak. For irrigated broad area agriculture on the Riverine Plain in the Murray and Murrumbidgee valleys., NSW Agriculture, Orange, 1999.

Huisman, J. A., Breuer, L., Bormann, H., Bronstert, A., Croke, B. F. W., Frede, H.-G., Gräff, T., Hubrechts, L., Jakeman, A. J. and Kite, G.: Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM) III: Scenario analysis, Adv. Water Resour., 32(2), 159–170, 2009.

Jeffrey, S. J., Carter, J. O., Moodie, K. B. and Beswick, A. R.: Using spatial interpolation to construct a comprehensive archive of Australian climate data, Environ. Modell. Softw., 16(4), 309–330, 2001.

Jensen, M. E.: Water consumption by agricultural plants (Chapter 1), Kozlowski, T.T. (Ed.), Water Deficits and Plant Growth. Vol. II. Plant Water Consumption and Response, Academic Press, New York & London, pp. 1–22, 1968.

Jensen, M. E., Burman, R. D. and Allen, R. G.: Evapotranspiration and irrigation water requirements, Amer. Soc. Civil Eng. Manuals and Reports on Engineering Practice No. 70, ASCE, New York, USA p. 332 , 1990.

Kirby, M., Mainuddin, M., Gao, L., Connor, J. D. and Ahmad, M. D.: Integrated, dynamic economic–hydrology model of the Murray-Darling Basin, in 2012 Conference (56th), February 7-10, 2012, No. 124487, Freemantle, Australia, Australian Agricultural and Resource Economics Society., 2012.

Kite, G. W. and Droogers, P.: Comparing evapotranspiration estimates from satellites, hydrological models and field data, J. Hydrol., 229(1), 3–18, 2000.

Ko, J., Piccinni, G., Marek, T. and Howell, T.: Determination of growth-stage-specific crop coefficients (Kc) of cotton and wheat, Agr. Water Manage., 96(12), 1691–1697, 2009.

Lascano, R. J.: A general system to measure and calculate daily crop water use, Agron. J., 92(5), 821–832, 2000.

Lascano, R. J., Van Bavel, C. H. M. and Evett, S. R.: A field test of recursive calculation of crop evapotranspiration, Trans. ASABE, 53(4), 1117–1126, 2010.

Lascano, R. J. and Van Bavel, C. H.: Explicit and recursive calculation of potential and actual evapotranspiration, Agron. J., 99(2), 585–590, 2007.

Lathuillière, M. J., Johnson, M. S. and Donner, S. D.: Water use by terrestrial ecosystems: temporal variability in rainforest and agricultural contributions to evapotranspiration in Mato Grosso, Brazil, Environ. Res. Lett., 7(2), 024024, 2012. doi:10.1088/1748-9326/7/2/024024

Liu, J., Wiberg, D., Zehnder, A. J. and Yang, H.: Modeling the role of irrigation in winter wheat yield, crop water productivity, and production in China, Irrig. Sci., 26(1), 21–33, 2007.

McMahon, T. A., Peel, M. C., Lowe, L., Srikanthan, R. and McVicar, T. R.: Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis, Hydrol. Earth Syst. Sci., 17, 1331–1369, 2013.

Mekonnen, M. M. and Hoekstra, A. Y.: The green, blue and grey water footprint of crops and derived crop products, Hydrol. Earth Syst. Sci., 15, 1577–1600, 2011.

Meyer, W. S.: Standard reference evaporation calculation for inland, south eastern Australia, CSIRO Land and Water., 1999.

Monteith, J. L.: Evaporation and environment, in Symp. Soc. Exp. Biol, vol. 19, p. 4., 1965.

Morton, F. I.: Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology, J. Hydrol., 66(1), 1–76, 1983.

Multsch, S., Al-Rumaikhani, Y. A., Frede, H.-G. and Breuer, L.: A Site-sPecific Agricultural water Requirement and footprint Estimator (SPARE: WATER 1.0), Geosci. Model Dev., 6(4), 1043–1059, 2013.

Oki, T. and Kanae, S.: Global hydrological cycles and world water resources, Science, 313(5790), 1068–1072, 2006.

Penman, H. L.: Natural evaporation from open water, bare soil and grass, Proceedings of the Royal Society of London. Series A. Math. Phys. Sci., 193(1032), 120–145, 1948.

Pereira, L. S., Paredes, P., Cholpankulov, E. D., Inchenkova, O. P., Teodoro, P. R. and Horst, M. G.: Irrigation scheduling strategies for cotton to cope with water scarcity in the Fergana Valley, Central Asia, Agr. Water Manage., 96(5), 723–735, 2009.

Priestley, C. H. B. and Taylor, R. J.: On the assessment of surface heat flux and evaporation using large-scale parameters, Mon. Weather Rev., 100(2), 81–92, 1972.

Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L. and Vanrolleghem, P. A.: Uncertainty in the environmental modelling process–a framework and guidance, Environ. Modell. Softw., 22(11), 1543–1556, 2007.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M. and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour. Res., 46(5), W05521, 2010. DOI: 10.1029/2009WR008328

Sheffield, J., Wood, E. F. and Roderick, M. L.: Little change in global drought over the past 60 years, Nature, 491(7424), 435–438, 2012.

Siebert, S. and Döll, P.: Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation, J. Hydrol., 384(3), 198–217, 2010.

Da Silva, V. de P., da Silva, B. B., Albuquerque, W. G., Borges, C. J., de Sousa, I. F. and Neto, J. D.: Crop coefficient, water requirements, yield and water use efficiency of sugarcane growth in Brazil, Agr. Water Manage., 128, 102–109, 2013.

Smith, M.: CROPWAT: A computer program for irrigation planning and management, Food and Agriculture Organization of the United Nations, Rome., Rome, Italy., 1992.

Steduto, P., Hsiao, T. C., Raes, D. and Fereres, E.: AquaCrop—The FAO crop model to simulate yield response to water: I. Concepts and underlying principles, Agron. J., 101(3), 426–437, 2009.

Sun, S., Wu, P., Wang, Y., Zhao, X., Liu, J. and Zhang, X.: The impacts of interannual climate variability and agricultural inputs on water footprint of crop production in an irrigation district of China, Sci. Total Environ., 444, 498–507, 2013.

Tabari, H.: Evaluation of reference crop evapotranspiration equations in various climates, Water Resour. Manag., 24(10), 2311–2337, 2010.

Tabari, H., Grismer, M. E. and Trajkovic, S.: Comparative analysis of 31 reference evapotranspiration methods under humid conditions, Irrigation Sci., 31(2), 107–117, 2013.

Toews, M. W. and Allen, D. M.: Simulated response of groundwater to predicted recharge in a semi-arid region using a scenario of modelled climate change, Environ. Res. Lett., 4(3), 035003, 2009. doi:10.1088/1748-9326/4/3/035003

Turc, L.: Estimation of irrigation water requirements, potential evapotranspiration: a simple climatic formula evolved up to date, Ann. Agron., 12(1), 13–49, 1961.

Viney, N. R., Bormann, H., Breuer, L., Bronstert, A., Croke, B. F. W., Frede, H., Gräff, T., Hubrechts, L., Huisman, J. A. and Jakeman, A. J.: Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) II: Ensemble combinations and predictions, Adv. Water Resour., 32(2), 147–158, 2009.

Vrugt, J. A., Ter Braak, C. J., Gupta, H. V. and Robinson, B. A.: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, Stoch. Env. Res. Risk A, 23(7), 1011–1026, 2009.

Wada Y., van Beek L. P., van Kempen C. M., Reckman J. W., Vasak S. and Bierkens M. F.: Global depletion of groundwater resources , Geophys. Res. Lett, 37(20), L20402, 2010. DOI: 10.1029/2010GL044571

Wada, Y., Wisser, D., Eisner, S., Flörke, M., Gerten, D., Haddeland, I., Hanasaki, N., Masaki, Y., Portmann, F. T. and Stacke, T.: Multimodel projections and uncertainties of irrigation water demand under climate change, Geophys. Res. Lett., 40(17), 4626–4632, 2013.

Wada, Y. and Bierkens, M. F. P.: Sustainability of global water use: past reconstruction and future projections, Environ. Res. Lett., 9(10), 104003, 2014. doi:10.1088/1748-9326/9/10/104003

Wang, H., Zhang, L., Dawes, W. R. and Liu, C.: Improving water use efficiency of irrigated crops in the North China Plain—measurements and modelling, Agr. Water Manage., 48(2), 151–167, 2001.

Williams, J.R., Jones, C.A., Kiniry, J.R. and Spanel, D.A., 1989: The EPIC crop growth model, Trans. ASAE, 32, 497-511.

Ye, B., Yang, D. and Ma, L.: Effect of precipitation bias correction on water budget calculation in Upper Yellow River, China, Environ. Res. Lett., 7(2), 025201, 2012. doi:10.1088/1748-9326/7/2/025201

Table 1. The five crop parameter sets for Kc.

Name (Reference)	Spatial reference	Kc_{ini}	Kc_{mid}	Kc_{end}
FAO56 (Allen et al., 1998)	Global	0.7	1.15	0.25
Harris (Harris, 2002)	Queensland	0.3	1.15	0.25
Kirby (Kirby et al., 2012)	Murray-Darling Basin	0.4	1.15	0.4
Meyer (Meyer, 1999)	Griffith, MDB	0.4	1.05	0.5
Hughes (Hughes, 1999)	Murray and Murrumbidgee valleys	0.3	1.0	0.6

Table 2. The six equations applied for the calculation of reference evapotranspiration.

Method	Abbreviation	Equation
FAO-56 Penman-Monteith (Allen et al., 1998)	PM56	$ETo_{PM56} = \frac{0.408 \cdot \Delta \cdot (\mathrm{R_n} - \mathrm{G}) + \gamma \cdot \frac{900}{\mathrm{T_{mean}} + 273} \cdot u_2 \cdot (e_s - e_a)}{\Delta + \gamma \cdot (1 + 0.34 \cdot u_2)}$
Priestley-Taylor (Priestley and Taylor, 1972)	PT	$ETo_{PT} = \alpha \cdot \left[\frac{\Delta}{\Delta + \gamma} \right] \cdot \frac{(R_n - G)}{\lambda}$
Hargreaves-Samani (Hargreaves and Samani, 1985)	HS	$ETo_{HS} = 0.0023 \cdot (\mathrm{T_{mean}} + 17.8) \cdot (\mathrm{T_{max}} - \mathrm{T_{min}})^{0.5} \cdot \mathrm{R_a} \cdot 0.408$
Turc (Allen, 2003; Turc, 1961)	TURC	$ETo_{TURC} = \alpha_T \cdot \frac{T_{mean}}{T_{mean} + 15} \cdot \frac{23.8856 \cdot R_s + 50}{\lambda}$
Areal – PET (Morton, 1983)	APET	$ETo_{APET} = b_1 + b_2 \left(\frac{1 + \gamma \cdot p}{\Delta} \right)^{-1} \cdot R_{TP}$
Point – PET (Morton, 1983)	PPET	$ETo_{PPETenergy-balance} = R_n - \lambda_P \cdot f_T \cdot (T_P - T_{mean})$ $ETo_{PPETvapor-transfer} = f_T \cdot (e_s - e_a)$

With ETo_{PM56}, ETo_{PT}, ETo_{HS}, PET_{TURC}, ETo_{APET}, $ETo_{Energy\text{-}Balance}$ and $ETo_{Vapor\text{-}Transfer}$ in [mm], extra-terrestrial radiation R_a, solar radiation R_s, net radiation R_n, soil heat flux density G and net radiation at equilibrium temperature R_{TP} in [MJ m^{-2}], equilibrium temperature T_P ,mean T_{mean}, minimum T_{min} and maximum T_{max} air temperature in [°C], wind speed u_2 at 2 m height [m s^{-1}], atmospheric pressure p, saturated e_s and actual e_a vapour pressure in [kPa], slope of vapour pressure curve Δ and the psychometric constant γ in [kPa °C^{-1}], latent heat of vaporization λ in [MJ kg^{-1}], and the dimensionless empirical constants b_1 and b_2 [-], the heat transfer coefficient λ_P [-], the vapour transfer coefficient f_T [-] and the humidity based value α_T.

Table 3. Average equally weighted irrigation water requirement ($\overline{IRR}$) [mm] during the growing season of wheat in all cells [n=3,969] of the MDB grouped by ET_o methods and K_c sets over the period 1986-2006 (APET: Areal potential evapotranspiration; PM56: FAO56 Penman Monteith; HS: Hargreaves-Samani; PPET: Point potential evapotranspiration; PT: Priestly-Taylor; TURC: Turc).

		Kc					
		Kirby	Hughes	Meyer	FAO56	Harris	$\overline{IRR}$
ET_o method	HS	381	381	372	349	336	364
	PT	661	671	654	618	580	637
	PPET	577	577	565	534	514	551
	PM56	365	362	355	344	324	350
	APET	357	354	347	329	315	340
	TURC	315	316	308	289	279	301
	$\overline{IRR}$	443	443	433	410	391	424

Table 4. Performance (R_B) and convergence (R_D) and reliability (R) coefficient of the ensemble members.

		FAO56	Harris	Hughes	Kirby	Meyer
	R_B	1	1	1	1	1
APET	R_D	1	1	1	1	1
	R	1	1	1	1	1
	R_B	0.96	0.96	0.96	0.96	0.96
PPET	R_D	0.99	1.00	0.99	0.99	0.99
	R	0.96	0.96	0.96	0.95	0.96
	R_B	1	1	1	1	1
HS	R_D	1	1	1	1	1
	R	1	1	1	1	1
	R_B	1	1	1	1	1
PM56	R_D	1	1	1	1	1
	R	1	1	1	1	1
	R_B	1	1	1	1	1
T	R_D	1	1	1	1	1
	R	1	1	1	1	1
	R_B	0.60	0.60	0.60	0.60	0.60
PT	R_D	0.98	1.00	0.85	0.88	0.90
	R	0.59	0.60	0.51	0.53	0.54

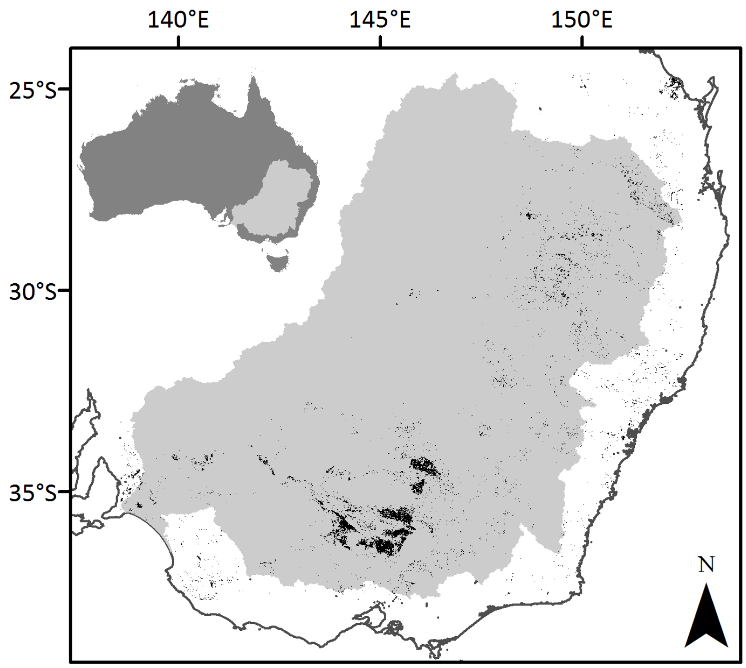

Figure 1. The Murray-Darling basin (MDB) is located in south-east Australia. Irrigated wheat areas (2005/06) across the MDB are indicated as black dots, n=3,969; cell size=1 x 1 km.

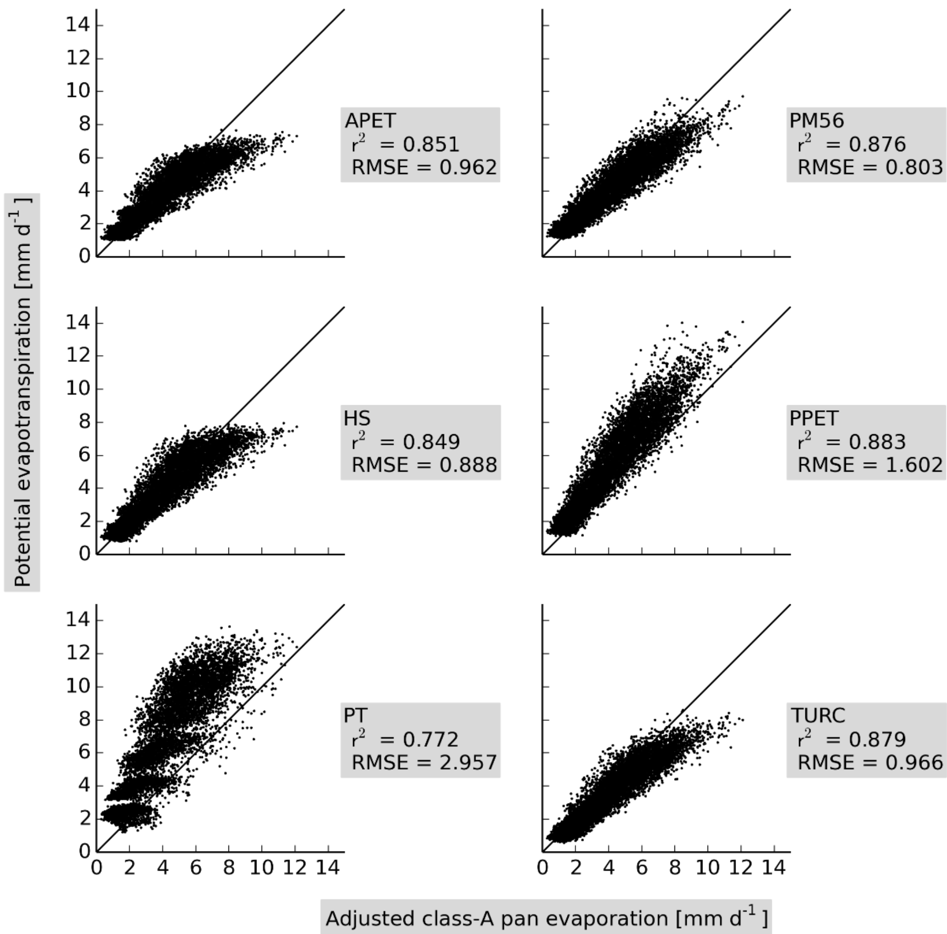

Figure 2. Comparison of daily measured class-A pan evaporation with simulated potential evapotranspiration at 34 sites in the MDB during the time period from 1986 to 2006. The class-A pan measurements have been adjusted with site-specific pan coefficients. The coefficient of determination (r^2) and the root mean square error (RMSE) are depicted for each ET_o method (APET: Areal potential evapotranspiration; PM56: FAO56 Penman-Monteith; HS: Hargreaves-Samani; PPET: Point potential evapotranspiration; PT: Priestly-Taylor; TURC: Turc).

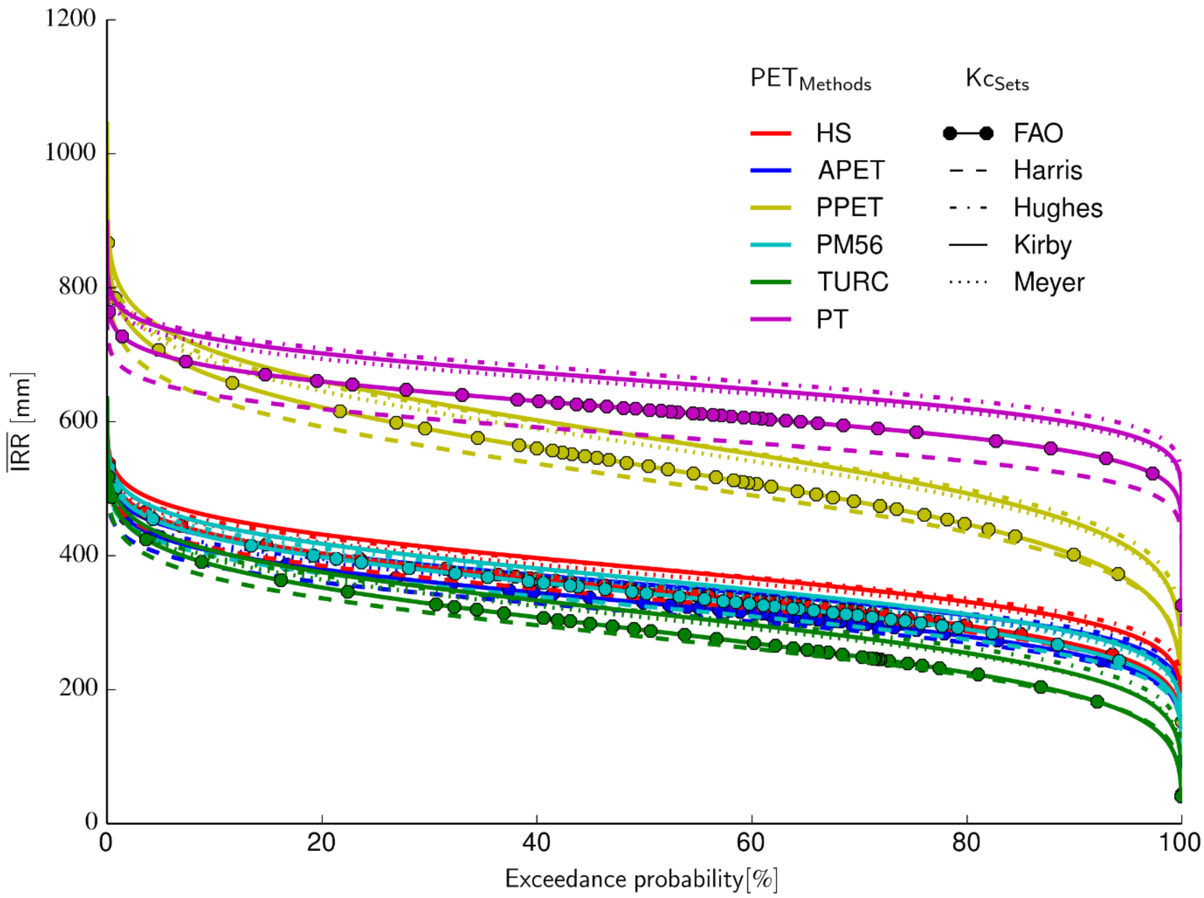

Figure 3 Exceedance probability of equally weighted average irrigation water requirement ($\overline{IRR}$) for wheat during the growing season. Averages have been calculated for each cropping area [n = 3969 = 100%] for the period 1986-2006. Colours indicate different ET_o methods (APET: Areal potential evapotranspiration; PM56: FAO56 Penman Monteith; HS: Hargreaves-Samani; PPET: Point potential evapotranspiration; PT: Priestly-Taylor; TURC: Turc) and symbols differentiate Kc sets.

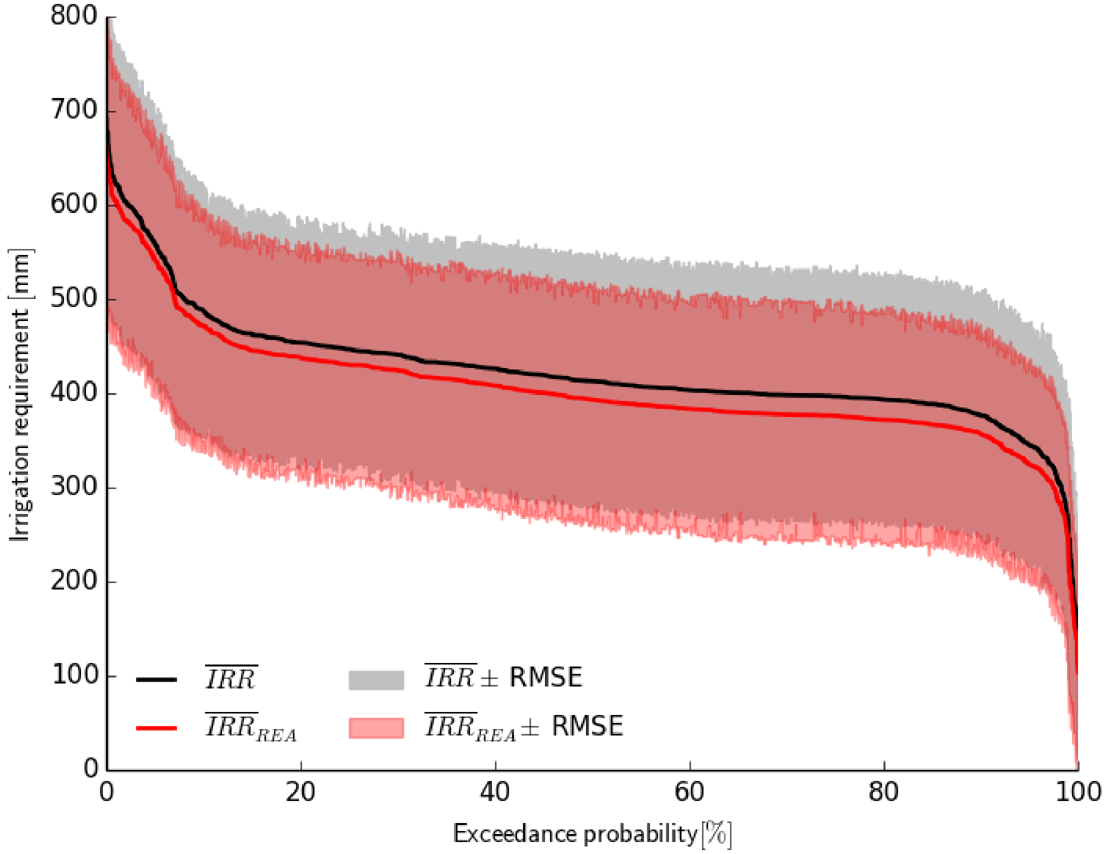

Figure 4 Cumulative density function of equally weighted ($\overline{IRR}$) and REA weighted ($\overline{IRR}_{REA}$) average irrigation water requirement for wheat during the growing season. Averages have been calculated for each cropping area [n = 3969 = 100%] for the period 1986-2006. Colours indicate the predicted root mean square difference (RMSE) of the ensemble of ET_o methods and Kc sets (APET: Areal potential evapotranspiration; PM56: FAO56 Penman Monteith; HS: Hargreaves-Samani; PPET: Point potential evapotranspiration; PT: Priestly-Taylor; TURC: Turc).

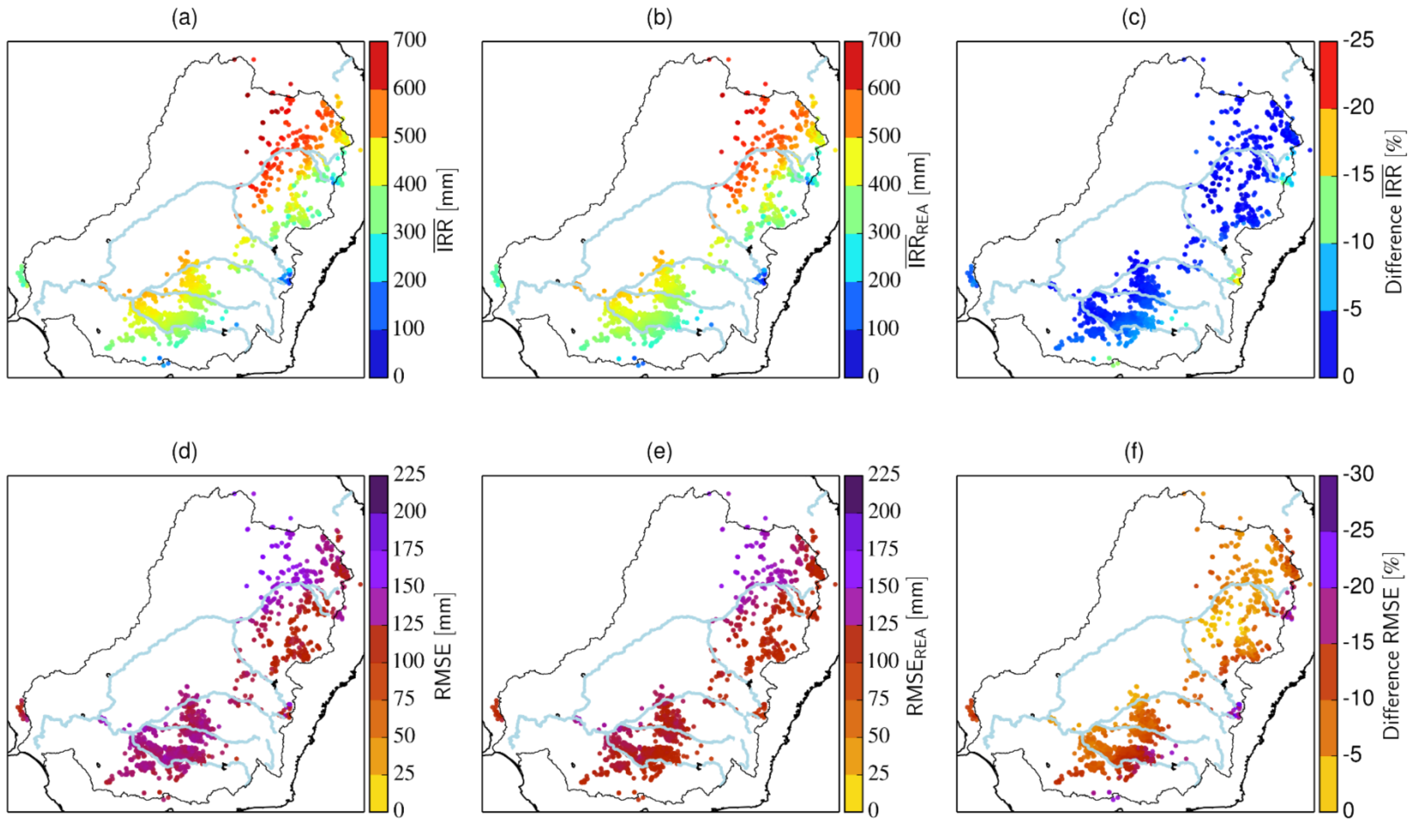

Figure 5. Average equally weighted (a) and REA weighted (b) irrigation water requirement during the growing season of wheat (1986-2006). Dots indicate irrigated cropping areas [n=3,969; cell size=1 x 1 km] (note: a buffer has been used to increase the visibility of the single grid cells). (c) illustrates the difference between both IRR calculations (b-a). (d) and (e) show the root mean square error between the 30 realizations and the equally weighted (d) and REA weighted (e) averages as well as the difference (f) between both calculations, respectively.