Introduction
Ocean salinity serves as a proxy of long-term shifts in Earth’s hydrological cycle, which has undergone regional intensification due to climate change. However, the historical record of salinity data is incomplete, as remote regions of the ocean went unsampled by research vessels, leading to large spatial sampling gaps. To overcome this, available in-situ measurements of ocean salinity collected since 1960 were gathered and the gaps in sampling were filled using a machine learning approach that creates fully interpolated salinity fields from the surface to 5500 m. The capability of this method is demonstrated by applying it to a historical run of the NASA-GISS climate model that has been decimated to the sample sparsity of actual observations. When comparing the filled-in salinity fields of NASA-GISS to the original run, good performance is achieved in capturing the annual changes in ocean salinity. Trends in both the deep ocean and individual basins are accurately estimated with minimal bias. Estimates of the annual ocean salt budget from the NASA-GISS model are presented here for the years 1960-2018.

Background
Global scale measurements indicate that the Earth has absorbed more energy than it has emitted over the past century (Hansen et al., 2005), a trend that has continued in recent decades (Trenberth et al. 2014). Warming due to this energy imbalance has been mitigated to a large degree by the ocean, which has gained roughly ten times as much heat over the past half century as all other parts of the Earth system combined (Church et al., 2011). As a byproduct, warmer oceans have intensified the global water cycle (Solomon et al. 2007), and compelling evidence of its modification over the past half century indicates an amplification of trends in evaporation and precipitation over the ocean (Durack et al. 2012; Skliris et al. 2014).

Currently, rainfall over the ocean is estimated to represent roughly 78% of the global total (Trenberth et al. 2007; Schanze et al. 2010), but this rainfall pattern is spatially heterogeneous, leading to regional contrasts that studies have shown to be escalating over multiple decades (Boyer 2005; Helm et al. 2010; Durack and Wijffels 2010; Skliris et al. 2014; Vinogradova and Ponte 2017). Given that ocean salinity measurements on multi-decadal time scales can act as a proxy for observed changes in evaporation and precipitation (Durack and Wijffels 2010; Helm et al. 2010; Durack et al. 2012; Terray et al. 2012; Pierce et al. 2012; Skliris et al. 2014; Lago et al. 2015; Vinogradova and Ponte 2017), developing good estimates of the change in multi-decadal global and basin scale ocean salt budgets is essential to understanding how the global hydrological cycle has evolved in the presence of anthropogenic warming.

A lack of direct observations of global precipitation and evaporation prior to satellite coverage has led to a reliance on coupled climate models (Stott et al. 2008; Terray et al. 2012; Pierce et al. 2012) or reconstructions using land-based instruments (Chen et al. 2002). These methods are thus tied to properties of the underlying models, including large uncertainties in fresh-water fluxes used as a boundary condition (Yu and Weller 2007; Schanze et al. 2010; Josey et al. 2013; Skliris et al. 2014) or their reliance on satellite derived basis functions (Chen et al. 2002) that are likely temporally variable. Alternatively, salinity has been used as a proxy for investigating shifts in the hydrological cycle, but due to data limitations past studies have been restricted to inferring linear trends between 5-10 year periods (Boyer 2005; Helm et al. 2010).

Data and Methods
In-situ salinity measurements taken between 1960-2018 are gathered from the World Ocean Database (WOD) 2013 (Boyer et al., 2013) (https://www.nodc.noaa.gov/OC5/SELECT/dbsearch/dbsearch.html, accessed 7-26-19). These are provided as separate datasets for the different instrument types, which in our case are the two bathythermograph datasets (XBT and MBT) that will be corrected, and the Ocean Station (OSD) and Conductivity-Temperature-Depth (CTD) which we use as reference data (Fig. 2, Step 1). We exclude casts that have been flagged by WOD for quality control issues. These measurements were then binned to the World Ocean Atlas grid (1x1 degree) for each calendar month and year of sampling.

Salinity fields from a historical run of the NASA-GISS CMIP5 model were interpolated to the World Ocean Atlas grid and points where no actual observations existed were excluded. Thus, the NASA-GISS model was decimated to the same sampling density as real-world observations. Realistic noise fields, representing the uncertainty in the observations from the limited spatial sampling were added to the NASA-GISS field as well. Subtracting out a monthly climatology produced anomaly fields of the NASA-GISS salinity that could then be gap-filled. Finally, these anomalies were averaged annually and an ensemble of neural networks was used to fill in the gaps for each year of salinity anomalies between 1960-2018.

Results

In order to determine the quality of the gap-filled product, the original NASA-GISS model run is compared here to the estimate obtained from running an ensemble of artificial neural networks. Trends in the gap-filled product recreate the original NASA-GISS quite well, even where data is most sparse. Individual ocean basins (Fig. 1), including those in the Southern Hemisphere, have been accurately recreated with minimal bias within model uncertainty.

![Graphs showing trends in salinity anomalies for different ocean basins](image)

**Fig. 1** Comparison of the trends in the gap-filled salinity anomalies (black) with 1-stddev uncertainty and the original NASA-GISS model run (red). Volume weighted averages for 700-2000 m for a. the Atlantic, b. the Pacific, c. the Indian, and d. the Southern Ocean basins.

Conclusion
A preliminary result demonstrating the ability of this method to fill gaps in the historical salinity record indicates that a climate model run with realistic noise can be accurately recreated with an ensemble of artificial neural networks. Trends in areas where observations are most sparse, such as the deep Southern Ocean, are captured well, indicating that this method can be applied to actual observations. Improved estimates of ocean basin-scale changes will better inform our understanding of the regional trends in the global water cycle and the amplifying of differences between relatively fresh and salty regions (Durack and Wijffels 2010; Trenberth 2011; Skliris et al. 2014; Lago et al. 2015), as well as the increased divergence in the salinities of the Pacific and Atlantic Oceans (Helm et al. 2010; Durack and Wijffels 2010).

References


